Understanding Children's Conceptual Development Through the Lens of the Representational Redescription Model

CHEUNG, Chi Ngai

A Thesis Submitted in Partial Fulfilment
of the Requirement for the Degree of
Master of Philosophy
in
Education

September 2008

Thesis/Assessment Committee

Professor Ni Yu-Jing (Chair)
Professor Wong Wan-chi (Thesis Supervisor)
Professor Shiu Ling-po (Committee Member)
Professor Au Kit-fong (External Examiner)

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisor, Prof. Wong Wan-chi, who has provided me with immeasurably insightful advice and given me unfailing support in my two years of master's study. Thanks to her, I have achieved significant development in both intellectual and personal aspects of my life. In the process of preparing my thesis, Prof. Wong has given me the greatest degree of freedom to explore my true interest. Her thoughtful advice has helped me to broaden my horizons and further improve the quality of my academic work. In the face of difficulties, her warm encouragement has given me a great deal of strength. My gratitude is also extended to Prof. Ni Yu-jing and Prof. Shiu Ling-po. They have given me many useful suggestions and comments, which helped me to understand the strengths and weaknesses of my thesis, so that I could make improvements.

The help of Pui Ching Primary School is of tremendous importance to the current study. I truly appreciate the effort made by the school principal, Mr. Lee See Yuen, vice principal, Ms. Leung Kam Sheung, and primary school teacher, Ms. Pong Chi Kwan. Without their help, I would not have been able to conduct any empirical study in their school. I am also very thankful to all of the participants and their parents, both in the main study and the pilot test. I would also like to thank Ms. Szeto Ka Ling and Mr. Cheung Chi Hin for helping me with the pilot test.

I would like to acknowledge the kind support and assistance of Miss Shi Huiwen and Mr. Fung Chun-Hai. Miss Shi provided a great deal of encouragement and support to me. She carefully read through my proposal and thesis in a very short period of time, and her advice and suggestions were very useful. By taking up the role of second rater, Mr. Fung undertook the tedious task of going through the complicated classification instructions and many pages of transcripts. I am very thankful for his effort.

Abstract of thesis entitled: Understanding Children's Conceptual Development Through the Lens of the Representational Redescription Model

Submitted by Cheung Chi Ngai
for the degree of Master of Philosophy
at The Chinese University of Hong Kong (July 2008)

In the field of conceptual development, the importance of the explicit-implicit dimension has not received proper attention. This thesis utilized such a dimension to understand the conceptual development in the physics and probability domains, and this was achieved by applying the Representational Redescription model (RR model). Children belonging to groups of 4 to 5 year olds, 6 to 7 year olds, and 8 to 9 year olds were studied. Differences in performance between the implicit and explicit representation were found, with evidence of implicit knowledge being developed before explicit knowledge was provided.

Study One investigated children's conceptual development in the block-balancing task. It was found that the behavioural success in block-balancing was a prerequisite for the discovery and verbalization of a more advanced concept: the naïve version of the law of torque. Furthermore, this explicit concept served as the precondition for the occurrence of the behavioural pattern that reflected this concept. Study One also provided evidence for the existence of the implicit geometric-centre theory as well as empirical evidence for the existence of the level-E1 representation, which is the unique contribution of the RR model. Study

Two investigated children's development in a probability-estimation task. It was found that intuitive estimation, which is sustained by implicit representation, coexisted with the explicit representation, and this implicit knowledge was better developed than the explicit knowledge among the participants of this experiment. More advanced explanations were not found in the group of 4 to 5 year olds. Some children in the group of 6 to 7 year olds and all of the children in the group of 8 to 9 year olds used the half rule in their explanations. The most advanced type of explanation in this study involved the concept of division/fraction, which could only be found in the eldest age group. Both studies found that less advanced explanations were phased out only gradually after the discovery of the more advanced explanations. The two studies yielded different results for the question of whether more advanced explanations were first discovered in easier trials or more difficult trials.

The application of the RR model in the physics and probability domains has provided the meaningful findings that are described in this thesis. The empirical results lend solid support to the RR model, and provide food for thought in the refinement of its content. These results suggest that the RR model can be applied to the study of conceptual development in new domains in the future. The results also strongly suggest that performance on the behavioural and verbal explanatory levels should not be viewed as the same. This recognition should have important implications for any future experimental design. In the practical arena, this thesis shed light on the process of learning, teaching, and assessment. It was shown in this thesis that the verbalization of a correct concept does not mark the end of learning or development. A teacher's role, viewed in the light of the current findings, should not be limited to the presentation of explicit knowledge, but should also be extended to facilitate students' gaining a fuller insight into their own concepts and knowledge. Assessment of learning should also take account of the implicit and explicit levels of knowledge.

在概念發展的文獻中，內隱/外顯的進路一直未獲適當重視。本文嘗試應用此進路，透過表徵重述理論來理解物理和機會率兩個範疇的概念發展，研究對象爲 4-5 歲, 6-7 歲及 8-9 歲之兒童。研究發現內隱表徵與外顯表徵有表現上的差異，且有證據顯示內隱知識的發展早於外顯知識。

實驗一研究兒童的概念如何在木條平衡任務中發展。研究發現，成功平均木條的行爲，是發現或言說有關力矩的初步概念的必要條件。而有關力矩初步概念的外顯知識，則是相關行爲模式出現的條件。實驗一亦證明了內隱的幾何中間點理論的存在，亦爲 E1 表徵層的存在提供了實證。E1 表徵層是表徵重述理論的獨特貢獻。實驗二研究兒童在機會率估計任務中的發展。研究發現，直

覺估計這種應由內隱表徵支持的能力，可與外顯表徵並存。而在本實驗的研究對象中，這種內隱知識的發展比外顯知識更爲進步。四至五歲的兒童未能提供較進階的解釋。部份六至七歲以及全部八至九歲的兒童，均曾提及使用「一半」爲答案的定準。本實驗最高階的解釋類型涉及除法/分數的概念，只有最年長的一組兒童能夠提供此類解釋。實驗一與二均發現，較初階的解釋是在較高階的解釋出現後逐漸消失。至於較高階的解釋會先出現於較深還是較淺的題目，兩個實驗有不同發現。

本文將表徵重述理論應用於研究物理與機會率的概念發展，得到許多有價值的發現。實驗結果既爲理論提供了實證，亦有助理論更臻完善。此理論可用於研究其他概念的發展。本研究的結果顯示行爲與言語解釋兩者的表現並不等同，這對實驗設計而言有重要的意義。在實際應用方面，本研究爲教學與評核方法帶來啓示。研究顯示，能夠述說正確的概念，並不等於學習或發展已達終點。教師的角色應不止於傳授外顯知識，更包括協助學生盡量對自己的概念與知識有充分的洞察。進行評核時，應同時考慮內隱與外顯的知識。

CONTENTS

LIST OF TABLES AND FIGURES

Tables

Figures

CHAPTER ONE: INTRODUCTION

The beauty and sophistication of the world is definitely fascinating, but our concepts and knowledge about this world can be even more so. Throughout human history, the search for the origins of concepts has never ceased. People are curious about how humans change from infants, who seem to understand nothing, into adults, who possess a large body of complex knowledge. The most amazing thing is that many of these conceptual changes occur without conscious effort or awareness. Though we are the agents of our own concepts, we are puzzled by their existence.

Philosophers and psychologists have tried hard to understand where concepts come from and how changes occur (e.g., Fodor, 1975; Locke, 1975; Prinz, 2002). "Nature versus nurture" is a classic question in psychology. Some hold that concepts come from nature and so people are born with knowledge, some argue the opposite position that all knowledge is acquired, and some prefer a position midway between these two. This question is asking about the source of concepts. But the issue about the source is not the only question that concerns conceptual development. How development actually occurs is also an important question. Developmental psychologists have proposed numerous theories that try to explain what occurs in conceptual development (e.g., Carey, 1991; Case, 1985; Fischer, 1980; Gopnik & Meltzoff, 1997; Keil, 1989; Piaget, 1952, 1955; Vygotsky, 1934/1987). The contents of these theories are often very different; however, many of them share the same underlying dimensions.

There are different approaches to investigating conceptual development. This thesis focuses on an approach that has not received much attention in the literature: the explicit-implicit dimension. Though only a few studies have been designed to investigate conceptual development using this dimension (e.g., Alibali & Meadow, 1993; Clements & Perner, 1994), it has given momentum to a great deal of research in other areas such as memory and learning (e.g., Cohen, Eichenbaum, Deacedo, & Corkein, 1985; Glisky, Schacter, & Tulving, 1986; Reber, 1967; Squire & Frambach, 1990). In this study, I will consider conceptual development by applying a model that is built on the explicit-implicit dimension, namely the Representational Redescription model.

Karmiloff-Smith's (1986, 1992) Representational Redescription model (RR model) is one of the most influential modern theories of cognitive development (Krist, Horz, & Schönfeld, 2005). Based on this model, Karmiloff-Smith (1992) developed a developmental theory that avoided the extremes of both nativism and empiricism, and reconciled the conflicts between a strict modular view of the human mind and a domain-general view of development. The theory provided a third way that led out of these dichotomies. The RR model is also a framework that insightfully integrates views of cognitive psychology and developmental psychology, acting as a conceptual bridge between these two subfields in psychology. It is not surprising that the RR model is referred to in studies that use the paradigm in cognitive psychology (e.g., Meulemans, Van der Linden, & Perruchet, 1998). Nor is it surprising to find developmental researchers who use the RR model to integrate their theories with findings on implicit learning or implicit memory (e.g., Steffler, 2001). The model introduced a new dimension for understanding development, namely the explicit-implicit dimension, which encouraged later researchers to study the role of implicit understanding in the course of development (e.g., Alibali & Meadow, 1993; Clements & Perner, 1994; Siegler & Stern, 1998). The basic assumption of the RR model is that explicitness and implicitness do not represent a dichotomy. The RR model classified representation into four kinds: Implicit, Explicit 1, Explicit 2, and Explicit 3. This classification influenced successive researchers as they tried to construct their theories. For example, Dienes and Perner (1986) acknowledged that, "In our framework there is no simple dichotomy between implicit and explicit knowledge. This owes much to Karmiloff-Smith's (1986, 1992) insistence that the basic dichotomy should be embellished by further levels of explicitness" (p. 748).

1.1 Overarching questions of the thesis

When applying the RR model to understanding conceptual development, the following questions arise, which provide the main direction for the investigation throughout this study:

1. Can the distinction between explicit representation and implicit representation be observed in this study of conceptual development? Do they have different roles, functions, and characteristics? Can implicit representation exist without explicit representation?
2. What are the processes that take place in implicit representation and in explicit representation in the course of conceptual development? Is there a U-shaped performance curve in this developmental process?
3. What are the dynamics between explicit and implicit representation? Is implicit representation replaced when explicit representation occurs, or do they coexist and perform different functions?
4. Is this conceptual developmental process itself a domain-general process?

To answer these questions, two experimental studies were designed for two different domains. By comparing the developmental processes related to conceptual understanding in the domains of physics and probability, it is hoped that more insight can be gained, and that the patterns of conceptual development can be better understood.

1.2 Study One: The block-balancing experiment

Karmiloff-Smith (1992) pointed out that the RR model is a domain-general model. She agreed that knowledge content in each domain is different because there are unique innate constraints in different domains. The RR model describes the development that occurs in different domains; the developmental process described by the RR model is a domain-general process. To prove this, Karmiloff-Smith (1992) discussed developmental research in five different domains: language, physics, mathematics, theory of mind, and notation. In the three domains that are related to conceptual development, namely physics, mathematics, and theory of mind, one experiment in the physics domain stands out as the most representative and provides the strongest empirical support for the RR model. This is the block-balancing experiment.

In the block-balancing experiment (Karmiloff-Smith & Inhelder, 1974/75), children were required to balance symmetrical and asymmetrical blocks over a narrow support. (The details of this experiment are discussed later in this thesis.) The experiment provided empirical support for the existence of the Explicit 1 (E1) level; the representation of this level is the abstraction of representation in the implicit level. E1 representation is not conscious and is not available for verbalization. The E1 level is a unique feature in the RR model that plays an important role in breaking up the explicit-implicit dichotomy. In the experiment, it was found that the behaviour of 6-year-old children was unconsciously constrained by an implicit theory about balancing, but they were unaware of this theory and therefore unable to verbally report it. This matched the description of E1 representation. Another line of support provided by this experiment is the U-shaped performance curve. It was found that both 4- to 5-year-old and 8- to 9-year-old children could balance both symmetrical and asymmetrical blocks, but 6-year-old children failed to do so. The RR model predicts that a decline in performance may happen during the course of development because, after behavioural mastery is reached, there is a phase of development that focuses on internal representation; in this phase, development is not data driven, so external data may be disregarded temporarily and may result in a decline in performance.

Pine and Messer (1999, 2003) found that the implicit theory suggested by Karmiloff-Smith and Inhelder (1974/75) could be verbalized by children. If this implicit theory is available to conscious access and can be verbalized once it appears, then the block-balancing experiment fails to support the existence of E1 representation. If the implicit theory is at first unavailable to conscious access and verbalization, but exerts its influence by affecting where children place the blocks, then the geometric-centre theory is represented at level E1 at this time. Even though, after a period, this implicit theory becomes available to both conscious access and verbalization, the existence of E1 representation is still supported. However, such critical information about the timing of availability is unclear in Pine and Messer's (1999, 2003) research.

Concerning the U-shaped performance curve, Krist et al. (2005) tested the performance of 4- to 6-year-old and 8-year-old children in the block-balancing task. They failed to replicate the finding of a U-shaped performance curve across the age groups, and instead they found quasilinear improvement with age. On the other hand, Pine and Messer (2003) found a U-shaped performance curve in the course of children's conceptual development. The age of their participants was between 5 years 8 months and 6 years 1 month. Unlike Karmiloff-Smith and Inhelder's (1974/75)

U-shaped curve, Pine and Messer's U-shaped curve is within the age group but across the five days of the experiment.

Given the importance of the block-balancing experiment to the RR model and the conflicting findings in the literature, children's performance in the block-balancing task was investigated in this study. There are four major aims of the study. The first aim is to find out whether the difference between implicit and explicit representation can be observed, and whether the block-balancing paradigm provides empirical support for the E1 representation. The former issue is associated with the basic assumption of the RR model. The second issue concerns with whether an intermediate representation level that exists between the extremes of the explicit-implicit dimension. Children's behavioural performance and verbal explanation was monitored trial by trial, in order to check whether Karmiloff-Smith and Inhelder's (1974/1975) claim that children could possess implicit theories that influence their behaviour is correct or not. The second aim is to see whether the bottom-up direction of development suggested by Karmiloff-Smith (1992) really exists. The third aim of this study is to see whether there is an across age-group U-shaped curve. The fourth aim is to study the developmental pattern of conceptual understanding in a block-balancing task. Variations within an individual, between individuals, and between age groups were studied carefully.

1.3 Study Two: The probability-estimation experiment

The RR model is intended to be a domain-general model for development. To prove that this claim is valid, the model must be applied to domains that it has not been applied to previously. To the best of the author's knowledge, the RR model has not been applied to studying the conceptual development of probability. The second experiment of this study used the RR model to investigate children's conceptual development of probability, and to consider what insights the RR model brings to the field.

The experiment focused on examining the following predictions that were deduced from the RR model:

1. There are implicit and explicit levels of representation for the probability concept.
2. The probability concept can be acquired by bottom-up development.
   i) Children can have some degree of behavioural success supported by implicit representation before they have explicit knowledge of how to carry out a correct computational procedure.

ii) Even if children receive no explicit instruction about the concept of probability, they can create explicit knowledge on the basis of their implicit knowledge. This representational redescription process can be facilitated by repeated practice with feedback.

Piaget and Inhelder (1951/1975) were the pioneering researchers in the field of probability development, and their theory dominated the field for many years (Reyna & Brainerd, 1994). However, their theory did not include a distinction between explicit and implicit knowledge. Piaget assumed that the developmental direction of the probability concept should be top-down because an explicit understanding of the particular defining features of the probability concept is the prerequisite of further behavioural improvement. For instance, concrete operational children make probability judgments based on the absolute frequency of the intended event rather than the ratio between intended and unintended events, because their explicit knowledge of combinatoric system and proportions has not yet developed, and does not do so until they enter into the formal operational stage.

Early research using the choice paradigm seems to agree with these suggestions (Chapman, 1975; Ross & Hoemann, 1975). In the choice paradigm, usually there were two events, event A and event B, and children were asked to decide which event would have the greater probability of happening. For example, there were two red marbles and one blue marble in jar A, and six red marbles and two blue marbles in jar B. When concrete operational children were asked to choose which jar they wanted to take a marble from, if their aim was to get a blue marble they usually chose jar B because the number of blue marbles was greater.

However, later research by Acredolo, O'Connor, Banks, and Horobin (1989) used the functional measurement method and found there was no difference in performance between concrete operational children and formal operational children. A possible explanation is that their probability-estimations have reflected an influence from both intended events and unintended events. Acredolo et al. believed their experiment showed that children possess implicit knowledge about probability. However, it is difficult to consider the distinction between explicit and implicit; given that Acredolo et al. did not ask children to explain their answers, the critical issue of the level of explicitness of these children's concept of probability was left unexplored in their study. Therefore, in this study, a modified design was applied to the experiment of Acredolo et al. (1989), and participants were required to provide verbal explanations in order to check the level of explicitness of their conceptual understanding.

When studying the direction of acquisition of the probability concept, results of previous experiments were all generalized from blocks of experimental trials; the temporal resolution of these experiments was quite low. The experimenters thus failed to capture the variations between trials that were critical to understanding how development actually occurred. Taking into consideration of the above limitations, children's performance and verbal explanations were monitored trial by trial in Study Two, with a view to constructing a picture of how conceptual changes occurred in a higher temporal resolution.

CHAPTER TWO: LITERATURE REVIEW

2.1 Approaches to studying conceptual development

Conceptualizing is a salient feature of the human mind and is closely related to the intelligent behaviours of humans. It is hard to imagine what would happen to human mental life if conceptual functioning ceased to work. With concepts, human beings are not only passive recipients of sensory experiences initiated by the environment, but are also active agents who construct understanding and hypotheses about the outside world. By using their conceptual functions, human beings can build a large body of knowledge, which enables them to live a life that is better than that of other animals.

In modern psychology, finding answers to the question of how concepts develop has been approached in various ways. The most traditional approach involves the rivalry between nativism (e.g., Fodor, 1975) and empiricism (e.g., Jackendoff, 1989), and debates whether conceptual knowledge is entirely innately specified, or is the result of interaction with the environment through the senses. Another approach is to ask whether human cognitive development is domain specific (e.g., Chomsky, 1980; Fodor, 1975) or domain general (e.g., Piaget, 1955). Some researchers ask whether children are universal novices, because if children become expert in a particular field, like chess, their performance is similar to that of an adult. The question posed by this approach is whether development is the same as the development of expertise. Recently, the role of theories in conceptual development has attracted many researchers' attention (Carey, 1991; Gopnik & Meltzoff, 1997).

In this thesis, the RR model was used to understand how conceptual development occurs because this model tries to approach conceptual development in another way, by considering the explicit-implicit dimension.

In cognitive psychology, there are several subfields related to implicitness. Reber (1967) discovered that if participants were required to remember lists of symbol strings that were generated by finite-state grammar, participants gradually acquired the grammatical rules, even though they were not informed of the existence of those rules. The participants acquired the rules incidentally and could not report the exact content of the rules. But the participants could remember novel grammatical strings better than ungrammatical strings, and when they classified strings as grammatical or ungrammatical their performance was above chance. Dienes, Altmann, Kwan, and Goode (1995) required participants to perform a secondary task (random number generation) while they were judging whether the stimuli were grammatical. It was found that responses that were judged by participants themselves as "guessing" were unimpaired, but knowledge associated with confident responses was impaired. This showed that implicit knowledge, which is unavailable to conscious access and represents itself as "guess," is qualitatively different from explicit knowledge.

In the literature on memory, there is also a distinction between explicit memory and implicit memory (Schacter, 1999). The explicit form, or declarative memory, includes memory for facts and events; the implicit form, or nondeclarative memory, includes memory for skills and habits, priming, simple classical conditioning, and nonassociative learning (Squire, 1994). Declarative memory depends on brain structures and connections in the medial temporal lobe and the diecephalon. Damage to these structures results in amnesia related to the failure of declarative memory. However, nondeclarative memory is still normal in these patients. How this distinction of memory can be related to conceptual development will be further discussed in a later section of this thesis.

Implicit knowledge is also found in developmental literature. Clements and Perner (1994) investigated children's theory of mind using a false belief task.

Children were told a story about a protagonist who hid something in one location. After the protagonist left the room, the hidden object was moved to a new location and the protagonist did not know about this change. Children were asked where the protagonist would look for the object. Clements and Perner found that before providing a verbal answer, children aged 2 years 5 months to 2 years 10 months looked at the new location of the object. Children aged 2 years 11 months to 4 years 5 months looked at the empty location where the protagonist thought the object should be. However, within this group, only 45% of them could give a correct verbal answer. This showed that implicit knowledge may exist without children's awareness, so children's verbal reports may not be an accurate indicator of the conceptual knowledge they possess. Implicit knowledge exhibited in this experiment may be the precursor of correct explicit knowledge about a false belief.

Gestures can reveal implicit knowledge possessed by children. Alibali and Meadow (1993) asked children to solve addition or addition-plus-multiplication problems and to explain how they solved each problem. Children's verbal explanations were often accompanied by gestures. Alibali and Meadow identified six types of procedures and the corresponding gestures. They found children often conveyed one procedure in a verbal explanation and another procedure by gestures at the same time; these children were classified as "discordant." Alibali and Meadow believed that the discordant state is transitional, because it predicts receptivity to instruction, and such a discordant state is both preceded and followed by a concordant state.

The empirical findings indicate that trying to understand conceptual development from the explicit-implicit representation seems to be a fruitful approach.

2.1.1 Direction of development and its implications for the study of conceptual development

If implicit and explicit representations of knowledge are different, their first occurrence may be different in the process of development. Karmiloff-Smith (1992) said the following:

> Development and learning, then, seem to take two complementary directions. On the one hand, they involve the gradual process of proceduralization (that is, rendering behavior more automatic and less accessible). On the other hand, they involve a process of "explicitation" and increasing accessibility (that is,

representing explicitly information that is implicit in the procedural representations sustaining the structure of behavior). (p. 17)

2.1.1.1 Top-down: The Adaptive Control of Thought model (The ACT model)

The top-down direction of development, that is, the process of knowledge represented in the explicit level guiding the development of the implicit level, can be explained by the ACT production model of skill acquisition (Adaptive Control of Thoughts; Anderson, 1982). This model proposes that the acquisition of cognitive skills involves two stages: the declarative stage and the procedural stage. In the declarative stage, people learn and memorize declarative knowledge about the specific domain that can be consciously recalled. During this stage, the execution is slow and resource demanding. Because declarative information needs to be retrieved from long-term memory, interpretation is required to turn declarative knowledge into an operation. This can place a heavy burden on the working memory. Anderson (1982) studied participants' performance in learning new geometry postulates and applying them to solve geometric problems. He found that, at the beginning, most of the participants' errors and their lack of speed could be attributed to the limitations of working memory.

The second stage of acquisition is called the procedural stage. Procedures are built to perform specific tasks, so that less time and working memory is required. Anderson (1982) observed that after solving four similar geometric problems, the protocol of a participant showed a qualitative change in the concepts involved. Since there was no verbal rehearsal of the statements of the postulate when solving the problem, this was evidence of reducing dependence on declarative knowledge. Also, a number of working-memory failures in the first protocol disappeared this time. Furthermore, in the first protocol, the application of the postulate was piecemeal, because the participant needed to identify and process every element in the postulate. But the postulate seemed to be matched in a single step this time. The process of declarative knowledge turning into procedures is called compilation, which can be subdivided into two stages: composition and proceduralization. Composition involves collapsing a sequence of productions into a single production, and proceduralization involves embedding factual knowledge into productions. When proceduralized, tasks can be performed efficiently and automatically, the procedure itself becomes encapsulated, and the intermediate steps of the task become unavailable to conscious access.

2.1.1.2 Bottom-up: The Representational Redescription model (The RR model)

The bottom-up direction of development, that is, turning knowledge embedded in an implicit level into explicit conceptual knowledge, can be explained by Karmiloff-Smith's RR model (1986, 1992). Karmiloff-Smith suggested that there are at least four levels of knowledge: Implicit (I), Explicit-1 (E1), Explicit-2 (E2), and Explicit-3 (E3). Knowledge of level-I is encoded in a procedural form and the content is sequentially specified. Since level-I knowledge is bracketed, it is available to other operators only as a whole and its individual components cannot be singled out for sharing with other procedures. Therefore, new knowledge is not integrated with existing level-I knowledge; rather, these knowledge types are stored separately. Intra-domain or inter-domain representational links cannot be established for this kind of implicit knowledge representation, so knowledge represented at this level is relatively inflexible.

Knowledge represented at level-E1 is not bracketed, so it can be available for potential intra-domain and inter-domain representation links, and is relatively more flexible than knowledge represented in level-I. However, an E1 representation is not necessarily available to conscious access and verbal reports. Level-E1 is generated by the representation redescription of level-I representation. Redescriptions are abstractions in higher level language, which enable higher level functions, like understanding the analogical relation between a zebra and a zebra crossing, but this is achieved at the expense of losing the detail of procedurally encoded information.

Knowledge represented at level-E2 is available to conscious access, but not to verbal reports. For level-E3, knowledge is represented in a cross-system code, which can be easily translated into natural language. Karmiloff-Smith believed conceptual knowledge represented in linguistic form, like knowledge printed in books, is directly stored in E3 in the process of learning.

Karmiloff-Smith (1986, 1992) insisted that behavioural mastery is not the end of development. She proposed that there are three phases of development that would occur iteratively in the development of each subdomain. In phase 1, learning is data driven and a new representation is stored as an adjunct to existing representations. This phase aims at reaching behavioural mastery. In phase 2, system-internal dynamics take over, and external data are temporarily disregarded, which may lead to a decline in performance. In phase 3, internal and external representations are reconciled, leading to improved performance. Unlike the stage model in which a child can only be in one stage at a time, the RR model allows a child to be simultaneously in phase 1 in one subdomain and in phase 3 in another subdomain.

A review of development (or psychological processes) from the top-down and the bottom-up directions provokes us to think about the study of conceptual development, and, in particular, about the values of the bottom-up direction as postulated by the RR model.

In the following section, I will continue to review the RR model in detail.

2.2 The Representational Redescription model

In this section, I will review the theoretical development of the Representational Redescription (RR) model, followed by a focused discussion of Karmiloff-Smith's (1992) book *Beyond Modularity*. I will first review Karmiloff-Smith's general view about human cognitive development, and then discuss the role of the RR model in development with respect to the domains of physics and mathematics. Finally, I will discuss an important concept in the RR model, namely "development beyond behavioural mastery."

2.2.1 Theoretical development of the RR model

The RR model was first proposed in a paper by Karmiloff-Smith (1986). The paper outlined three phases of development, and introduced four different levels of development along the explicit-implicit dimension (i.e., I, E1, E2, and E3; see section 2.1.1.2). The data that provided support for the RR model in that paper came from the linguistic domain. In her book, *Beyond Modularity*, Karmiloff-Smith (1992) discussed how the RR model can be used in other domains like physics, mathematics, theory of mind, and notation. When applying her model to different developmental data in the book, she did not distinguish between level-E2 and level-E3, and called them collectively E2/3, because she said there was no direct evidence to prove the existence of E2, but she stressed that she was open to the possibility of consciously accessible spatial, kinaesthetic, and other non-linguistically-encoded representation. She also loosened the sequential constraints of the implicit level in the domain of drawing in response to the findings of other researchers.

The RR model proposed by Karmiloff-Smith (1986, 1992) was an attempt to explain the emergence of conscious access to knowledge, and to understand how representation becomes more manipulable and flexible. The core of the RR model is the representational redescription process. Karmiloff-Smith (1992, p. 18) explained that the "representational redescription process is a process by which implicit information *in* the mind subsequently becomes explicit knowledge *to* the mind."

2.2.2 The RR model in *Beyond Modularity*

Often when other researchers mention the RR model, only the three phases of development and the different levels of explicitness are discussed. In the book *Beyond Modularity*, the RR model is not merely a theoretical tool for understanding development along the explicit-implicit dimension, but it is also the third way out of the dichotomy, which is generated by understanding development as completely domain-specific or completely domain-general. Karmiloff-Smith (1992) saw the nativist view and Piaget's constructivist view as not necessarily incompatible. She believed that infants are born with domain-specific constraints that can either be detailed specifications or nonspecific, like attention bias or skeletal outline. Whether the constraints are specific or nonspecific depends on the domain. These constraints enable learning to take place by trimming down the number of possible hypotheses, because infants must process a tremendous amount of information. For nonspecific constraints to develop into knowledge, inputs from the physical environment or the sociocultural environment are essential. These inputs interact with inborn constraints, and structure the mind of infants and children.

Karmiloff-Smith (1992) suggested that development involves two directions. One direction is gradual modularization; a process in which knowledge becomes more encapsulated and less accessible to other systems. The other direction is redescription; a process in which knowledge becomes progressively more accessible. She saw the representation redescription process as a bridge between implicit inborn constraints and explicit theories that can be consciously constructed, tested, modified, and explained.

In the following sections, I will first review Karmiloff-Smith's argument as presented in *Beyond Modularity*, and then discuss how she applied her RR model to the domains of physics and mathematics, because these two domains are the ones most closely related to conceptual development.

2.2.2.1 The role of representational redescription in development: The physics domain

Karmiloff-Smith (1992) quoted Spelke's (1990) studies to illustrate that very early in their lives, infants possess the principles of perceptual processing, like boundedness, cohesion, rigidity, and no action at a distance. This knowledge appears much earlier than Piaget suggested, but infants do not have all of the knowledge from the outset. For example, Spelke, Breinlinger, Cacomber, and Jacobson (1992) found that 6-month-old infants showed surprise when they saw a situation in which the law of gravity was violated, but 4-month-old infants did not show any surprise. Karmiloff-Smith agreed that infants possess a rich, coherent, and stable representation of these principles, but she denied that very young infants have theories. Theories are represented at level-E1 or higher. The principles possessed by infants are implicit and cannot be used for purposes other than the fixed input-output relation. She named this kind of implicit representation as level-I representation, and suggested that it exists in the form of procedures for responding to environmental stimuli. To transform implicit representations into explicit linguistically coded theory, representation redescription is required. She provided three reasons to support the idea that a representational redescription process is necessary: while 4-month-old infants show surprise when seeing one object pass through another, 2-year-old children fail to give an explicit response other than a habituation response. Secondly, children's explicit theories usually strongly resemble constraints that control earlier behaviour. Third, there are examples of theory-in-action—children's behaviours demonstrate that they possess a particular theory, but they cannot encode that linguistically and they may even not be aware of that theory. However, their persistent behavioural tendency proves the existence of an implicit theory.

Karmiloff-Smith used the following example to explain her last two reasons. Baillargeon and Hanko-Summers (1990) found that 7- to 9-month-old infants looked at a symmetrical object for significantly longer when the object exhibited an impossible support relation (upper right of Figure 2.1), but they were not surprised when asymmetrical objects exhibited an impossible support relation. Karmiloff-Smith and Inhelder (1974-75) studied how 4- to 9-year-old children tried to balance blocks on a narrow support (Figure 2.2). They found that the 6 year olds stubbornly tried to balance all of the stimuli blocks at the geometric centre, regardless of whether the blocks were evenly weighted or not. This demonstrated their implicit geometric-centre theory—all objects should be balanced symmetrically along their length. The theory is called implicit because the 6-year-old children could not verbally report this belief. Moreover, this geometric-centre theory resembles that of the 7- to 9-month-old infants, in the sense that the minds of both the children and the infants are constrained by symmetry, as shown in their responses.

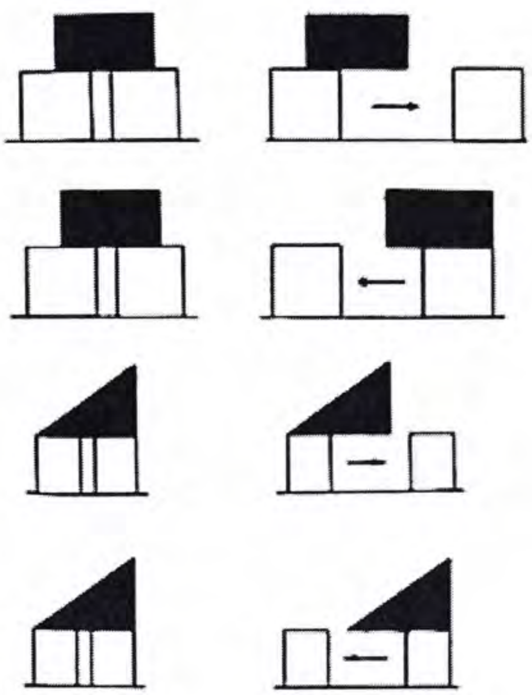

Fig. 2.1 Possible and impossible support relations (From Baillargeon & Kanko-Summener, 1990)

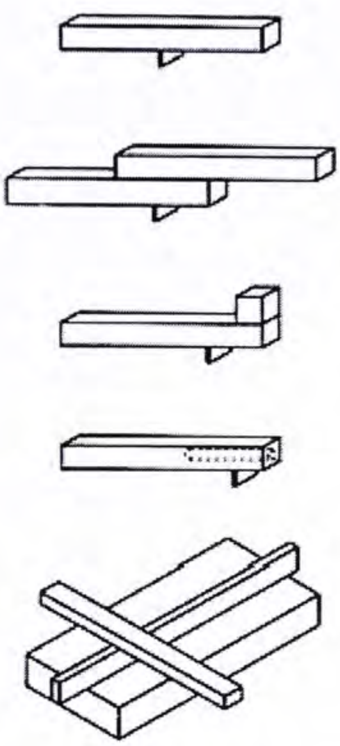

Fig. 2.2 Stimuli for the block-balancing task (From Karmiloff-Smith, 1992, p. 85)

To summarize, Karmiloff-Smith (1992) used findings in the physics domain to demonstrate that infants possess rich knowledge, but the knowledge is represented in an implicit form. Representational redescription is required to develop explicit theories bases on implicit knowledge.

2.2.2.2 The role of representational redescription in development: The domain of mathematics

Karmiloff-Smith (1992) quoted Antell and Keating (1983) to point out that even neonates can detect number difference. She agreed that some of the constraints on counting may be inborn. Gelman and Gallistel (1978) identified the following constraints on counting: one-to-one correspondence, stable ordering, item indifference, order indifference, and cardinality. Karmiloff-Smith believed that one-to-one correspondence may be operative in a neonate's and young infant's discrimination of arrays; although toddlers make mistakes in counting, they rarely violate the one-to-one correspondence constraint. However, Karmiloff-Smith disagreed with Gelman and Gallistel that the principal of cardinality is innately specified. She quoted Gelman and Meck's (1986) study, which showed that when 2 year olds were asked to give the cardinal value of a set, they correctly counted and repeated the last number. However, even though the same set of objects was used for every trial, the 2 year olds recounted on every new trial. Karmiloff-Smith attributed this to the reason that knowledge embedded in the counting procedure is not manipulated as separate components by the children, so they must reenact the whole procedure in every trial. She believed that the counting procedure itself does not equate to understanding cardinality because implicit principles that are embedded in procedures are not directly available to children. But with stable procedural representation accompanied by behavioural mastery, representation redescription can take place. Knowledge embedded in the procedure can be abstracted, redescribed, and represented in another format that is different from the procedural encoding. The implicit cardinality principle embedded in the counting procedure can be converted into explicit knowledge.

To summarize, Karmiloff-Smith (1992) used findings in the domain of mathematics to demonstrate how representation redescription could turn implicit knowledge that is embedded in procedural representation into explicit knowledge, and how procedural representation in the brain could act as an internal source of concepts.

2.2.2.3 Important concept of the RR model: Development beyond behavioural mastery

In developmental research, usually researchers are contented when they see the behavioural results of children reach peak performance for a particular task, and this point marks the end of the research because there is no more development to study. Karmiloff-Smith (1992) insisted that behavioural mastery is not the end of development, it is only the end of phase 1 development, which is primarily data driven and makes alterations in response to negative feedback. In phase 1, the object of the analysis is the external environment and the aim is to build up an internal representation of the external environment in order to achieve behavioural mastery.

In phase 2, the object for the analysis is internal representation and the aim is to disembed the knowledge that is implicit in procedures that are stored in the mind. A temporary drop in performance is very often caused by awareness of information that is already embedded in procedures that have been carried out successfully. This awareness may lead to overemphasis on that information, but this overemphasis is corrected in phase 3. Therefore, a U-shape may be found in the overall trend of development.

For example, in the notation domain, Karmiloff-Smith (1979a) designed a map-drawing task for children. Participants needed to invent notation for recording the correct route between two places. Along the route, there were many bifurcations, and the participants needed to record which branch was the correct one.

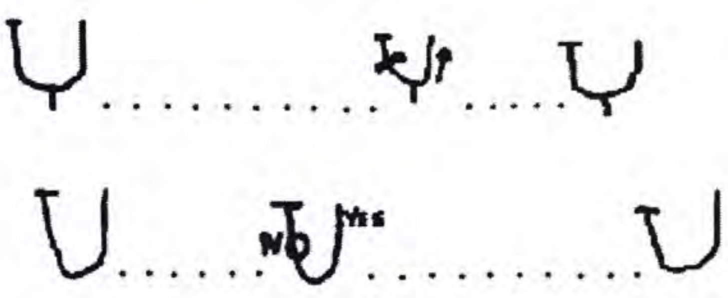

Fig. 2.3 Change of notation (From Karmiloff-Smith 1992, p. 153)

The microdevelopmental changes of participants showed that development did not stop at behavioural mastery. After the children had used a form of notation that was sufficient for recording the route information, they suddenly introduced redundant information to their notation. As shown in Figure 2.3, one of the participants added an arrow, and another added "yes" and "no," but such additions do not convey any additional information. This kind of change was not found in the first six records of bifurcations, so Karmiloff-Smith believed that it meant behavioural success, and the stable representation behind the behavioural success was the prerequisite for such a change. After a few trials, participants discarded the redundant information and returned to their original, more economic notation.

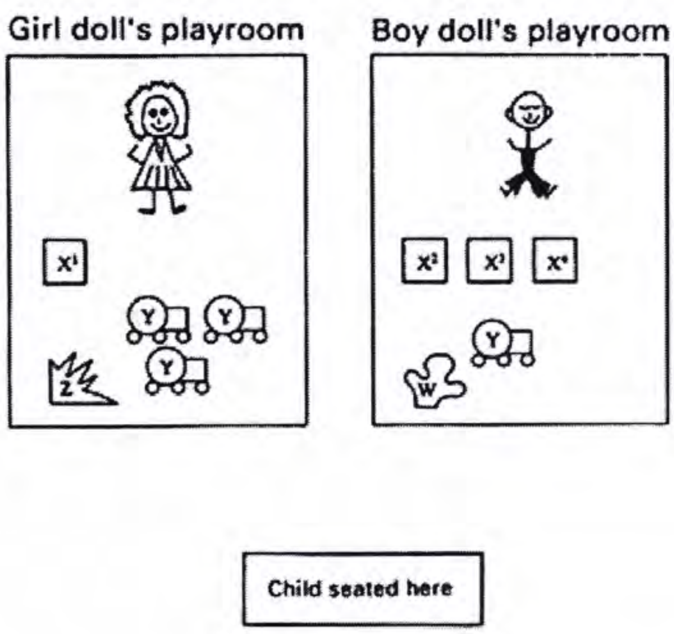

Fig. 2.4 The experimental context for the article task (Karmiloff-Smith, 1992, p. 56)

In the linguistic domain, Karmiloff-Smith (1979b) investigated the development of the article system in children who spoke French. In French, the indefinite article "un" (or feminine counterpart "une") has two different meanings: it can act as an indefinite reference like the English "a," and it can have a numerical function like the English "one." In the experimental context illustrated in Figure 2.4, there are several Y (cars) in the girl doll's room, but there is only one Y in the boy doll's room. When the experimenter said, "Lend me the car," the participant could infer from the article "the" that the experimenter was talking about the boy doll's car, since there was only one car in that room. If the experimenter said, "Lend me a car," the participant could infer from "a" that the experimenter is talking about the girl doll's car, because there were several cars in that room. It was found that at age 3 years, children are able to understand the difference in meaning between the definite and the indefinite article. In other words, they had reached behavioural mastery. However, around the age of 5 or 6 years, children regress behaviourally and in the study, they responded to the indefinite version of the statement by picking up the car from the doll who only had one car. Karmiloff-Smith interpreted this as a development of representation because the child interpreted the statement as "lend me one car" instead of "lend me a car." The child is now aware of both the indefinite function and the numeral function of the word "un."

2.2.3 Remarks on the RR model and the representational redescription process

Karmiloff-Smith (1992) stated that only the representation redescription process is the core of the RR model. It was possible that there may be more or fewer levels than she originally suggested, or the relationship between her four levels is not as she originally thought, she indicated that the RR process was not challenged by these modifications. Although the need for behavioural mastery or the three recurrent phases of development can be modified or refuted, the validity of the RR process is not necessarily affected. Karmiloff-Smith pointed out that the RR model only loses its plausibility when empirical data show that the representation redescription process does not exist.

After considering Karmiloff-Smith's (1992) clarification, I propose that the following three terms should be treated as different conceptual entities in the present study: Representational Redescription process, Representational Redescription model, and Representational Redescription theory. The representational redescription process is the process of creating a new level of representation based on the original representation. The RR model is the model built with the representation redescription process as the core, supplemented by the four hierarchical levels of representation, three phases of development, and the concept of development beyond behavioural mastery. The representational redescription theory is a hypothesis about human cognitive development which proposes that human cognitive ability originates from domain-specific constraints, and development requires interaction between environmental inputs and innate constraints. The representational redescription process involves development by creating a more explicit representation of the originally implicit knowledge, so as to enable another conceptual development process, like explicit theory testing, which involves both graduated modularization and demodularization processes.

2.2.4 Remarks on the meaning of the top-down and bottom-up directions of development

The terms "top-down" and "bottom-up" are commonly used in the literature of psychology. Generally speaking, "top-down" implies a process that is guided by more abstract, explicit, conceptual, and generalized content. "Bottom-up" usually implies a process that is guided by more concrete, implicit, perceptual, and local features. Bottom-up can also mean the process of generalization from instances, and top-down means the application of a general rule (Shiu & Sin, 2005).

In this thesis, the terms "top-down" and "bottom-up" are used to indicate the developmental directions along the explicit-implicit dimension. Applying these directions to the RR model, top-down or bottom-up development implies interrepresentational development. Top-down development refers to the development that starts from a more explicit level of representation, and results in the formation of a less explicit representation. Conversely, bottom-up development starts from a more implicit level of representation, and results in a less implicit level of representation.

In the following sections, I will review the empirical evidence for the RR model.

2.3 Empirical evidence related to the RR model in the linguistic domain

2.3.1 Support for the explicit-implicit distinction

Paradis (2004) proposed that there are two types of linguistic processes: one is explicit, consciously controlled, and requires effort; and the other is automatic, uncontrolled, and effortless. When one learns and applies explicit rules of language, like pedagogical grammars or various types of theoretical grammars, these kinds of processes rely on declarative knowledge and require conscious effort. With practice, learners become more fluent and proficient. However, this improvement is not the result of processing the explicit rules faster; instead, the conscious process is replaced by an automatic, procedural, and implicit computational mechanism. The process of learning explicit linguistic rules and the acquisition of implicit computational mechanisms are different processes. An explicit rule remains explicit and can be recalled like any other information stored in declarative memory, unless it is forgotten. Implicit computation is not automatization of an explicit rule; explicit knowledge does not evolve or change into implicit linguistic competence. No matter how the performance of the speaker resembles an explicit rule, the explicit rule remains an abstraction based on the outcome of language processing, and is not a correct descriptions of actual cerebral computational procedures. Explicit knowledge of language and linguistic competence rely on different memory systems that have separate, anatomically distinct, neural substrates. The function of explicit knowledge is to guide attention, and to provide a model for monitoring and judging the quality of automatically generated structures during practice.

People with genetic dysphasia learn their native language by learning explicit rules (Paradis & Gopnik, 1997). Normal 3-year-old children can automatically and unreflectively generate the plural of a nonword, like changing "wug" to "wugs." People with genetic dysphasia cannot do this at age 13 years, nor at age 43 years, nor at age 63 years, because they do not possess a generalized mechanism for processing plural forms. They learn plural forms as individual words and the plural forms are stored in their lexicon in the same way as nonimpaired speakers store irregular or idiosyncratic plurals.

Paradis (2004) pointed out that Alzheimer's disease damages explicit memory. He quoted a finding by Meguor et al. (2003) that Japanese-Portuguese bilingual patients who had Alzheimer's disease experienced greater difficulty with kanji in Japanese and with irregularly spelt words in Portuguese. Paradis (2004) explained this by the difference in the ability of the patients to learn words and to acquire syntax; most experimental primates have been able to learn an impressive number of words, but the best of them have been able to master only a very rudimentary syntax.

Further evidence that proves the dissociation between explicit and implicit language processing comes from people who have semantic dementia, which causes the explicit memory for word meaning to be impaired. These people have poor comprehension ability, perform poorly on the naming task, and often make semantic errors. However, their ability to process syntax is preserved. They can speak and write fluently, but commit regularization errors, such as their pronunciation of "have" rhymes with "gave," and words with irregular spelling and sound correspondence are regularized, such as spelling "caught" as "cort" (Snowden, Neary, & Mann, 1996).

2.3.2 Spelling development

Nunes, Bryant, and Bindman (1997) investigated children's development in spelling past tense verbs, in a longitudinal study over 3 years that included 363 children from three grade levels (Grades 2, 3, and 4). They found that children's development can be divided into five stages. In Stage 1, children have not yet established a stable letter-sound correspondence so they generate unsystematic spelling of word endings. In Stage 2, they begin to frequently transcribe word endings related to past tense; however, the ending is not necessarily spelt as “ed,” for example, “kissed” may be spelled as “kist.” In Stage 3, children begin to produce the “ed” spelling; however, they are insensitive to the meaning of “ed.” Thus, they overgeneralize “ed” to nonverbs and they also use “ed” for irregular verbs. In Stage 4, children no longer overgeneralize “ed” to nonverbs, but they still use “ed” for irregular verbs. In Stage 5, children finally master the use of “ed” and use it only for regular verbs. This developmental trend resembles the overgeneralization of the block-balancing task, in which children developed an implicit geometric-centre theory for block-balancing (Karmiloff-Smith & Inhelder, 1974/1975).

Steffler (2001) argued that the RR model can be used as a framework that provides better understanding of the development of the ability to spell. Critten, Pine, and Steffler (2007) tried to apply the RR model to study spelling of past tense words. They asked children to spell words and to judge whether stimuli words were correctly spelled. They found that a group of children failed to provide an explanation for their answers, and their responses included statements such as, “I don’t know,” or “It looks right.” They could not make use of their own level-I representation, which would have helped them to make a judgment, to spell words, and to provide a verbal explanation. This agrees with previous neuronal data that suggests that such processing is supported by an implicit process, and also agrees with the RR model's prediction that level-I representation must go through the representational redescription process before being available to conscious access and verbal reports.

2.4 Empirical evidence related to the RR model in the field of strategy development

2.4.1 Siegler's research on strategy: behavioural changes precede insights

Siegler and Jenkins (1989) provided evidence that conceptual change does not necessarily come from conscious inference, or by revising an explicitly represented hypothesis. Processes that are unreachable for consciousness also play an important part in conceptual change. Siegler and Jenkins investigated changes in children's strategy for addition; the focus of study was on how children change from using "sum strategy" to "min strategy." For example, to calculate 3+5, children using sum strategy put up three fingers first, then put up five fingers, and finally counted from one, saying "1, 2, 3, 4, 5, 6, 7, 8" to obtain the answer. If the min strategy was used, the child counted from the larger addend, regardless of its position in the question. So, the child said "5, 6, 7, 8" or "6, 7, 8," perhaps simultaneously putting up one finger on each count beyond 5. Siegler and Jenkins adopted a microgenetic approach to create a high-resolution picture of how changes occurred. Children were invited to explain their method of obtaining the answer after each question. It was found that children's insights about the better strategy that they had just used fell along a continuum.

Siegler and Jenkins classified the children's explanation into two categories. Children who counted unambiguously from the larger addend and accurately described their way of doing it were classified as meeting the strict criterion. For those counting from the larger addend but who were unable to explain what they did, or those who used an approach that combined the characteristics of the min strategy with other strategies were classified as meeting the loose criterion. Many children failed to acknowledge their change in strategy, and even after the new strategy had been used several times, they could not provide an explanation for that strategy. For example, here is the protocol of one of the participants, Whitney, when she first discovered the min strategy:

> E: How much is 4 + 3?
>
> W: (mumble) 6, 7, I think it's 7.
>
> E: Seven, OK, how did you know that?
>
> W: Because I'm smart and I just knew it.
>
> E: Can you tell me, I heard you counting. I heard you. Tell me how you counted.
>
> W: I just—I didn't count anything … I just added numbers onto it.
>
> E: Can you tell me how you added numbers?

W: No.

E: Come on Whitney—come on, we have to do this, OK?

W: OK, 3, add one makes 4, add one more makes 5, add one more makes 6, add one more makes 7, add one more makes 8.

E: Wait, but how did you know what 4 + 3 was?

W: Cause I did what I just showed you. I just used my mouth to figure it out.

(Siegler & Jenkins, 1989, p. 86-87)

Siegler and Jenkins explained that although Whitney said she counted from 3, the videotape showed Whitney's initial mumbled sound was clearly a single word, and sounded most like 5. In later trials when her counting could be heard clearly, Whitney consistently counted from the larger addend, but she often denied she had counted. Another participant explained to the experimenter that the answer "just popped into my head" and in two of three trials she used min strategy.

If an improvement of strategy resulted from conscious inference, insights should not fall along a continuum, and changes in later use of a strategy should be clear cut, because once a better strategy is discovered, the better strategy should be used in the rest of the test. In the studies, most of the children fell back to an inefficient strategy after initial discovery of an efficient strategy. Siegler and Jenkins found that the explicitness of the explanation predicts subsequent generalization of the strategy to other problems.

Another discovery of the study is that there is a bridging strategy called a "short-cut sum" between "sum strategy" and "min strategy." When using this strategy to calculate "3 + 5," children say "1, 2, 3, 4, 5, 6, 7, 8" and may put up one finger on each count. To change from the sum strategy to the short-cut sum strategy, one discovery is needed: the numerical value of the addend is equal to the number itself. This knowledge is embedded in the procedure of sum strategy. Throughout the experiment, there was no direct instruction about this knowledge. Therefore, the representational redescription process should have taken place to disembed this information and contributed to the development.

2.4.2 Dixon's research of the representational redescription process

Another line of research conducted by Dixon (Dixon & Bangert, 2002; Dixon & Dohn, 2003; Dixon & Kelley, 2007) focused on the role of redescription in the process of representational change. It should be noted that the meaning of "redescription" in Dixon's papers had a slightly different meaning when compared to

Karmiloff-Smith (1992) and this thesis. Dixon's line of research mainly focused on one of the functions of redescription: to disembed knowledge from procedures. The other meaning of redescription, the process of a concept developing from one representation level to another level, was excluded. This difference was explained in the following footnote in Dixon & Bangert (2002, p. 919): "As Karmiloff-Smith noted [1992, p. 23], the appropriateness of the hierarchical model and the process of representational redescription are separate issues. The current article focuses on the process of representational change." Dixon's meaning of "representational change" refers to addition to or change in conceptual content rather than a change at the level of representation.

Dixon and Bangert (2002) studied whether theory revision or the representational redescription process is responsible for a representational change and they found that both methods contribute to representational change. They used the gear-system task to investigate the problem. Figure 2.5 illustrates an example of the question used in the gear task. The gear with a single arrow on it is the driving gear that provides the force to move all the other gears in the system. The gear with a shelf holding a small pile of coal is the fuel gear. Participants needed to decide whether the fuel gear would turn clockwise or counter-clockwise, and to place the train that needs coal on the appropriate side.

Fig. 2.5 Example of the gear problem

(Dixon & Bangert, 2002, p. 920)

One way to solve the problem is to simulate the movement of the gears, tracing how the force transfers from gear to gear. This strategy is called the "Figure 8" strategy because the path results from the force tracing movement resembles the line that is drawn for a figure of eight. A more advanced strategy is to represent the gear system as an alternating sequence, in which a clockwise gear is followed by a counter-clockwise gear, and no simulation is need. This strategy is known as the left-right (LR) strategy. The knowledge of the alternating sequence is embedded in the "Figure 8" strategy, because "clockwise, counter-clockwise" is repeated in the procedure of tracing the movement. Therefore, the LR strategy can be acquired by the representational redescription process from the "Figure 8" strategy. The alternating sequence can be abstracted further. If the total number of gears is even, the final gear turns the same direction as a driving gear, and, if the number of gears is odd, the final gear goes in the opposite direction. This strategy is called the "counting parity" strategy. Dixon and Bangert constructed a model using the five immediate trials before the discovery of the LR strategy, and they found that the probability of discovering the LR strategy is significantly predicted by an increase of accuracy in recent trials. This agrees with the prediction based on the RR model that positive feedback is important for development beyond behavioural mastery. On the other hand, although theory revision predicts representational change when current representation produces error, it is not responsible for discovering the LR strategy. However, Dixon & Bangert found that low accuracy preceded the discovery of the "Figure 8" strategy. Therefore, Dixon & Bangert believed that the representational redescription process and theory revision play a different role in representational change. Dixon and Kelley (2007) hypothesized that theory revision is responsible for the search for the best relational structure in a child's repertoire that minimizes errors. The representational redescription process detects regularities in the interaction between the external environment and the internal representation, then produces a new representation of the relational structure. This new relational structure is added to a child's repertoire of relational structures, which is then available for the creation of a hypothesis in future, and becomes a source for theory revision.

Dixon and Dohn (2003) compared the effect of transfer between two kinds of acquisition methods: knowledge that is disembedded using the representational redescription process, and knowledge that is obtained from direct instruction. They designed two structurally similar tasks, one is the “Balance Beam” task, and the other is the gear task mentioned before. Participants attempt the balance-beam task before they try the gear task. Figure 2.6 shows an example stimulus for the balance-beam task. Three balance beams are connected by flexible joints, which are represented as ovals in the figure. The arrow indicates the point where the beam would be pressed down, and participants were required to answer whether the right-most arm would move up or down.

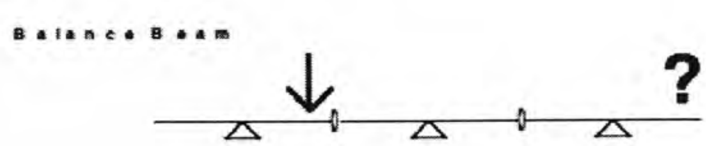

Fig. 2.6 The balance beam task

(Dixon & Dohn 2003, p. 1084)

Dixon & Dohn identified three kinds of strategies for the task: force tracing, up-down, and counting parity. Participants who use the force tracing strategy simulate the movement of beams by moving their hand in a continuous sinusoidal motion. The up-down strategy assigns “up” and “down” labels to beams alternately, so that the up beam is followed by the down beam, and the down beam is followed by the up beam. The process for the counting parity strategy is to count the number of beams, and to decide the answer depending on whether the total number of beams is an odd or an even number. The force tracing strategy is the easiest strategy to acquire and most closely resembles the gear task situation. The knowledge about the up-down strategy is embedded in the force tracing strategy. Participants were divided into two groups; one group received no training, so they had to solve the problem by themselves. The other group received training about the up-down strategy. Dixon & Dohn predicted that the representational redescription process would produce a more abstract representation of the alternating sequence relationship, and thus would facilitate the use of the left-right strategy in the gear task. The results confirmed their prediction. In the gear task, the untrained group discovered the left-right strategy earlier than the trained group. Moreover, in the 10 trials that immediately followed the first discovery of the left-right strategy, the untrained group fell back to the "Figure 8" strategy less frequently than the trained participants. This showed that knowledge acquired by the representational redescription process is qualitatively different from that acquired by direct instruction.

2.5 Empirical evidence related to the RR model in the physics domain:

The block-balancing problem

2.5.1 The original experiment

The block-balancing task, in which participants were requested to balance different kinds of blocks, was first introduced by Karmiloff-Smith and Inhelder (1974/1975). The stimuli that were used are illustrated in Figure 2.7. Among these stimuli, some of them balance at the geometric centre, and some of them do not. The experiment was divided into two phases. In phase I, the order of the blocks was not fixed, and the participants were free to choose which block they wanted to try first. About twelve months later, half of the participants in phase I were interviewed again.

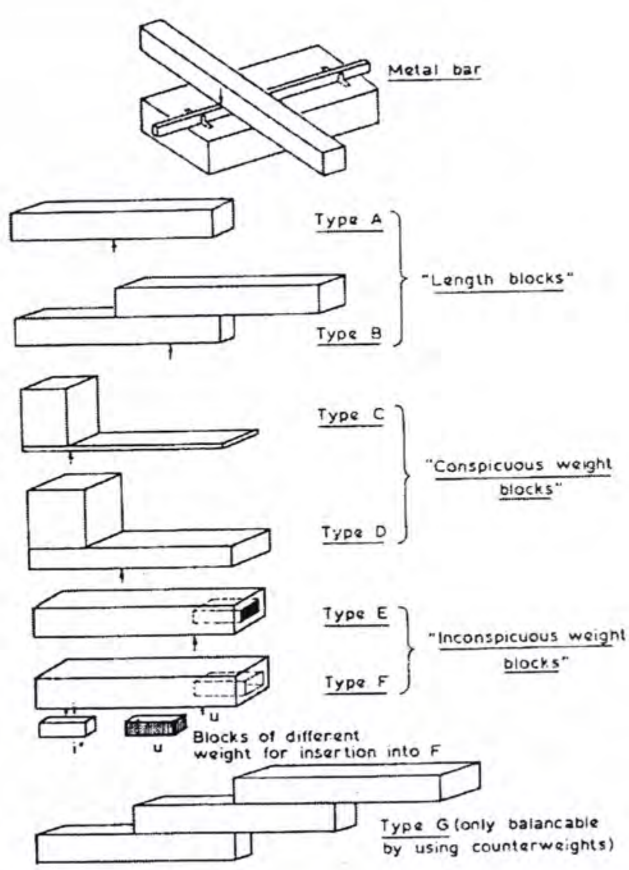

Fig. 2.7 Stimuli used in the Karmiloff-Smith & Inhelder (1974/75, p. 197) experiment

Karmiloff-Smith (1992) used the results of this experiment to provide empirical support for the RR model. Four year olds were able to balance these blocks by using proprioceptive feedback. They placed the block on the metal support and, by trial and error, they could eventually balance the blocks. Karmiloff-Smith interpreted the 4 year olds' performance as possessing knowledge at level-I. They failed to make use of the information that they had gained in previous successful trials. They did not pick a block that was identical to previous successful trials, and so they expressed no sign of an ability to use their successful experiences. Karmiloff-Smith observed that 6-year-old children put every block, regardless of whether the block was symmetrical or asymmetrical, at the geometric centre. When an asymmetrical block fell, the children put the block at the geometric centre again, until they could no longer accept the failure and claimed that it was impossible to balance the block. Karmiloff-Smith interpreted this as a sign of E1 representation. The geometric-centre theory is a reduced redescribed representation, a common feature that holds for many situations. Children aged 8 or 9 years could balance both symmetrical and asymmetrical blocks and gave relevant explanations. Karmiloff-Smith interpreted that they possessed E2/3 representation, which is an explicit knowledge about the geometric centre and a naïve version of the law of torque.

This developmental sequence fits nicely with the three recurrent phases of development proposed by Karmiloff-Smith (1992). In phase 1, learning is data driven and new representations are stored as adjunctions. This phase aims at reaching behavioural mastery. In phase 2, system-internal dynamics takes over, and external data is temporarily disregarded, which may lead to a decline in performance. In phase 3, the internal and the external representations are reconciled, leading to improvement in performance. This U-shaped performance curve was reflected in the previously mentioned block-balancing task.

To summarize, Karmiloff-Smith & Inhelder's (1974/1975) experiment provides support to the RR model in the following ways:

1. The 4 to 5 year olds treated each block as an individual problem and they did not use their successful experiences from previous trials. This is evidence for level-I performance, since each representation was added as an adjunction and no links were built between the new representation and the old representation.
2. The 6 year olds had the implicit geometric-centre theory, which provides support for the existence of level-E1. The geometric-centre theory also resembles an infant's ability to judge impossible balance relations. Infants showed surprise that

symmetrical objects showed an impossible balance relation, but they did not show any surprise about asymmetrical objects.

3. The 8 to 9 year olds could verbalize explanations and balance both types of block, 4 to 5 year olds could balance but cannot explain, and 6 year olds could not balance in certain cases. This U-shaped profile is predicted by the RR model.

2.5.2 Later research

When Krist, Horz, and Schönfeld (2005) reviewed the research that uses the block-balancing paradigm, which was after Karmiloff-Smith & Inhelder (1974/75), Krist et al. (2005, p. 185) said that "To the best of our knowledge, there is only one line of such research: a series of studies conducted by Pine and Messer." My findings are similar, except that I found one more journal paper written by Messer without Pine (Messer, Joiner, Light, & Littleton, 1998), and the experiment conducted by Krist et al.. I will first discuss the research that is most relevant to the issue of conceptual development in research conducted by Pine and Messer (Messer, Joiner, Light, & Littleton, 1998; Pine, Lufkin, & Messer, 2004; Pine & Messer, 1998, 1999, 2000, 2003).

When Messer, Joiner, Light, and Littleton (1998) studied the block-balancing task performance of 3- to 8-year-old children, they asked the children to balance eight blocks, half of which were symmetrical and half were asymmetrical. In Study One, they found that none of the children fulfilled their definition of Level I (the blocks were placed at the centre initially no more than twice, and eventually the child was able to balance at least six blocks). Messer et al. claimed that they could not provide a reason for this because the blocks used were very similar to those used by Karmiloff-Smith and Inhelder (1974/75). However, this is not true, Messer et al. did not use "invisible asymmetrical blocks" (Type E and F in Karmiloff-Smith & Inhelder's (1974/75) study) and the Type B symmetrical block was not used.

Different initial placements were made by 17% of the children, depending on the type of block, and Messer et al. classified these children as Level IV in their new seven-level scheme.

In Study Two, Messer et al. (1998) introduced a paper-and-pencil version and a computer version of the balancing-block task. For the paper-and-pencil task, a block was drawn on paper and children were required to indicate where the fulcrum should be placed to balance the block. For the computer task, children were required to move a cartoon character to the position that balanced the block. Feedback about success or failure was given and the children were allowed to attempt as many times as they liked. Messer et al. predicted that children would perform better in the physical version of the task because proprioceptive feedback could not be used in the other tasks. However, children appeared to perform better in the paper-and-pencil task, and this cannot be explained by the practice effect because the paper-and-pencil task was the first task that the children attempted. Of the children who were classified as Level II (initial central placement and failure to balance asymmetrical blocks) in the physical task, 29% were classified as Level IV in the paper-and-pencil task. Among the children classified at Level III in the block- balancing task (initial central placement and success in balancing both symmetrical and asymmetrical blocks), 55% of them performed at the same level in the paper-and-pencil task. Comparing with the computer task, among children classified as Level II in the physical task, 39% stayed at the same level in the computer task, and 29% of them performed at Level III in the computer task.

Messer et al. (1998) explained superior performance of the computer task by the practice effect, because the computer task was attempted after the physical task. This reveals a design fault of the experiment, because participants always attempted the paper-and-pencil task first, then the physical task, and attempted the computer task last. Since the aim of the experiment was to find out the performance difference between tasks, the task sequence should have been randomized among the participants to avoid practice effect. For the superior performance of the paper-and-pencil task, Messer et al. explained that placing the fulcrum under the block may reduce the salience of symmetry. I have concerns about their explanation, because it is difficult to understand why placing the fulcrum under the block may reduce the salience of symmetry, and how this reduction in salience of symmetry can affect performance. I suggest that the paper-and-pencil task and the computer task were superior to the physical task because the two tasks required the children to move the fulcrum but not the block. In Karmiloff-Smith and Inhelder's (1974/1975) experiment, children were likely to attribute their failure to properties related to the blocks, rather than the point of balance that they attempted. Therefore, if children were required to move the fulcrum rather then the block, this may have the effect of sensitizing them to the critical point of the problem, helping them to avoid distracting information about the blocks.

In both of the above experiments, a verbal explanation from children was not obtained. This is a serious deficit, because the RR model emphasizes development beyond behavioural mastery and development along the explicit-implicit dimension. Just observing success or failure without understanding the children's belief about the task is not adequate, because such kind of observation neither validates nor falsifies the RR model.

Pine and Messer (1999) used the same block-balancing paradigm and they asked the children for an explanation. They found that there were children who could fulfil their definition of Implicit level (balance at least 75% of both types of blocks, no systematic strategy about initial placement, and no verbal explanation), which contradicts Messer et al. (1998), because Messer et al. found no participant fitting their Level I definition. Pine and Messer (1999) mentioned that a typical child at this level, like Oliver (6 years 1 month), appeared to have no concept of balancing. When Oliver was asked for an explanation for his success in balancing an asymmetrical block, he replied, "I just put it on like that."

Pine and Messer (1999) found that 45% of the children who were classified as E1 were able to explain their geometric-centre theory, by referring to placing blocks "in the middle" or having "both sides equal." It seems that Pine and Messer misinterpreted Karmiloff-Smith's (1992) levels of representation, the RR model is not for classifying children's performance of a task, but is a framework for classifying representation. If geometric-centre theory can be consciously accessed and verbally reported by a child, then the representation of geometric-centre theory in that child should be classified as E3. In Study Two, Pine and Messer deleted level-E2 and inserted four new levels between level-I and level-E3; these are Implicit transition, Abstraction nonverbal, Abstraction verbal, and Explicit transition.

Levels	Definition
Implicit transition	Failed at least 3 of 4 attempts for each type of block, initial centre placement, no explanation nor relevant variables mentioned
Abstraction nonverbal	Only successful with symmetrical blocks, initial centre placement, no explanation nor relevant variables mentioned
Abstraction verbal	Only successful with symmetrical blocks, initial centre placement, explains centre strategy
Explicit transition	Success in 3 out of 4 attempts with both types of blocks, may have initial centre placement but rapid correction, explains strategy for balancing both types of blocks

Table 2.1 New levels introduced by Pine & Messer, 1999

In terms of classifying children's performance, these additional levels help to provide a clearer differentiation. However, these additions also further deviate from the original purpose of the classification system. These levels are not describing representation. Instead, they are describing children's block-balancing behaviour, which certainly violates Karmiloff-Smith's (1992) original intention of creating a domain-general framework for classifying representation. The modified framework is entirely domain specific.

In Study 2, using their modified framework, Pine and Messer (1999) found that the result of the British Picture Vocabulary Scale was not related to the level at which the children were classified, which means that successful verbalization of an explanation was not the result of more advanced language ability. Since their classification in this study so greatly modified the levels specified by Karmiloff-Smith's RR model, I have reservations about using this result to support using the RR model.

Nevertheless, their finding that 6- to 7-year-old children were able to verbalize geometric-centre theory deserves attention. Karmiloff-Smith (1992) pointed out that the implicit geometric-centre theory is represented at the E1 level in 6 year olds, this implicit theory should be unreachable by consciousness and cannot be verbalized. Karmiloff-Smith (1992) also pointed out that the explicit form of geometric-centre theory should be possessed by 8- to 9-year-old children. If 6 year olds can verbalize the geometric-centre theory, then the geometric-centre theory in their mind is not represented at level-E1, but at level-E3. If the geometric-centre theory is immediately represented at level-E3 when it appears, then the block-balancing experiment does not provide empirical evidence for the existence of level-E1 representation. Unfortunately, more detailed information about the children's verbalization of geometric-centre theory cannot be found in the Pine and Messer (1999) experiment, such as, whether there is a developmental sequence within individuals, or whether children first possess implicit geometric-centre theory, and later develop it into E3 representation.

Pine and Messer (2003) studied children's conceptual development of block balancing over a 5-day period, for participants aged between the ages of 5 years 8 months and 6 years 1 month. The experiment included a block-balancing task, a Predition task, and a period of free play time. They used the classification system introduced in Pine and Messer's study (1999), with an additional level, E4, which was defined as asymmetrical blocks being placed on or near the correct off-centre balance point. From the perspective of understanding conceptual development, this addition does not seem to be useful. Pine and Messer's E3 and E4 levels are both E3 representation according to Karmiloff-Smith's (1992) model, because both of them could be accessed by consciousness and could be verbally reported. Moreover, from the results of the experiment, the children who were placed at this level seemed to retrieve the answer from memory (the same stimuli were used repeatedly throughout the five days) rather than acquiring further conceptual change, because all E4 children reached this level on day 5, and the only difference between E3 and E4 was the initial placement position.

Pine and Messer's (2003) framework helped to summarize the children's performance in the course of development. Of all the transitions within these five days, 80% were moving to a higher level, 8% stayed in the same level and 12% slipped back. Among 25 participants, 7 children were first classified as performing at the implicit level (can balance at least two asymmetrical blocks), then went through the Abstraction nonverbal level or the Abstraction Verbal level (both require balancing no more than one asymmetrical block). For the predition task, children were invited each day to place blocks that they thought could not be balanced into one group, and blocks that could be balanced into the other group. The results of the children's predictions at different levels of representation are shown in Figure 2.8. We can observe a U-shaped curve that shows best performance in the prediction test at two ends of the explicit-implicit spectrum, namely the implicit level, the level of explicit transition, level-E-3, and level-E-4.

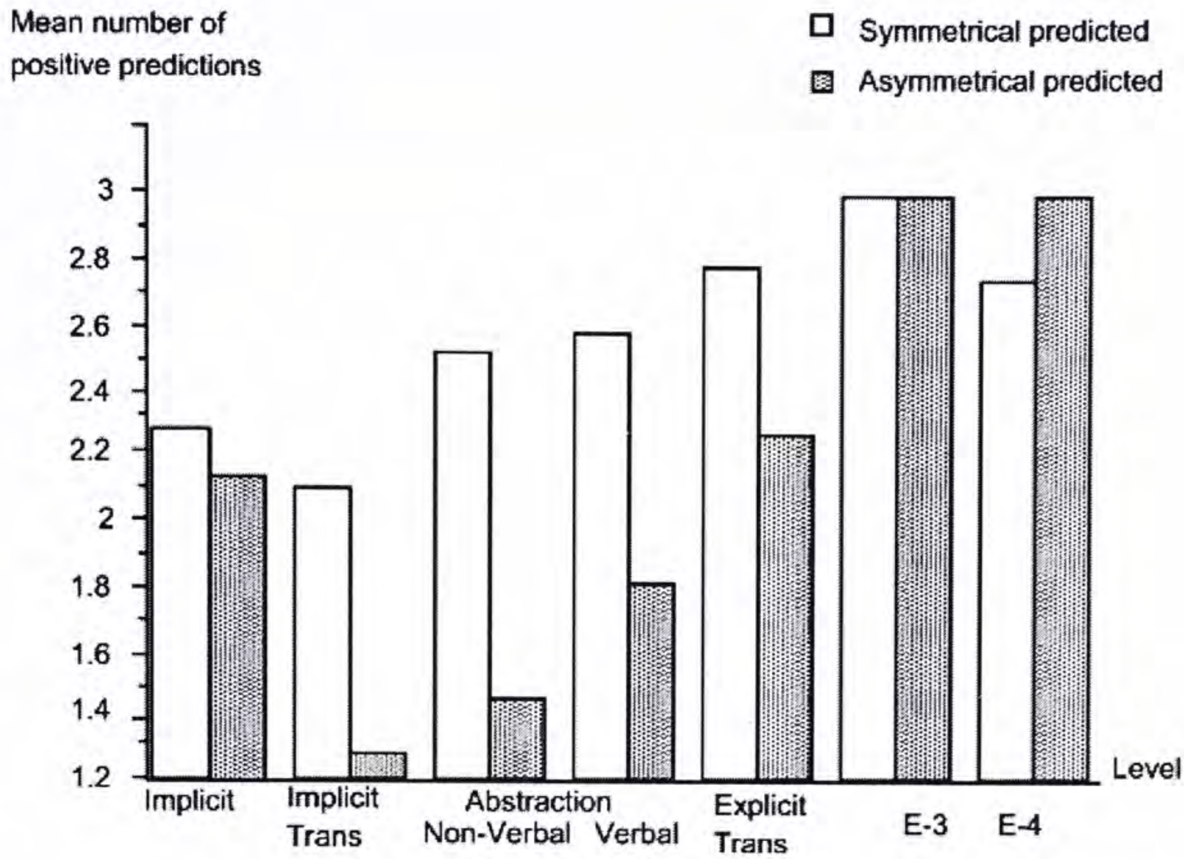

Fig. 2.8 Children's prediction of whether a certain block can be balanced (Pine & Messer 2003, p. 295)

When observing verbalization of the geometric-centre theory, among 25 children, 16 participants were classified as Abstraction verbal at least once. Among these 16 children, three of them were classified at this level on day 1. This suggested the possibility of the geometric-centre theory being available to conscious access and verbal reports earlier than Karmiloff-Smith and Inhelder (1974/75) suggested. However, it should be noted that in the experiment conducted by Pine and Messer (2003), the levels were assigned based on performance in the block-balancing task. The block-balancing task was preceded by the picture-story task and the Predition task. In the picture-story task, participants were required to choose a picture that depicted a correct balance point. In the prediction task, participants were required to group blocks that could be balanced into one group and blocks that could not be balanced into another group. It may be the case that for the three participants who were classified as Abstraction verbal on day 1, their geometric-centre theory was implicit at the beginning of the experiment. After the practice provided by the two preceding tasks, the implicit geometric-centre theory became available to conscious in the block-balancing task. It is important to know whether geometric-centre theory is available to conscious access and verbal reports immediately after it exists, or whether the theory takes time to develop from unconscious theory-in-action into conscious theory. If the previous situation is true, then the block-balancing experiment does not provide evidence for the existence of E1 representation. If the later proposition is true, then the block-balancing experiment still provides good empirical support for E1 representation.

Krist, Horz, and Schönfeld (2005) divided the evidence supporting the RR model into two groups. The evidence in the stronger group is concerned with development beyond behavioural mastery, especially the U-shaped curve. They believed that if the RR model was correct, then the following outcomes should be found: the rate of success and relative frequency of correct adjustments should follow a U-shaped performance; and response time and frequency of initial midpoint placement of an asymmetrical block should follow an inverted U-shaped performance. They studied 4-, 5-, 6- and 8-year-old children. The children were asked to balance five blocks in the order A, B, C, D, E (shown in Figure 2.9) in two series of trials. No verbal explanation was required. The success scores showed a significant age effect, but no U-shaped curve was found. The age effect was not found in the Midpoint scores, which measure the number of trials in which a block was first placed at its geometric centre, but interaction between age and symmetry was significant. No U-shaped curve was found for the Midpoint score either.

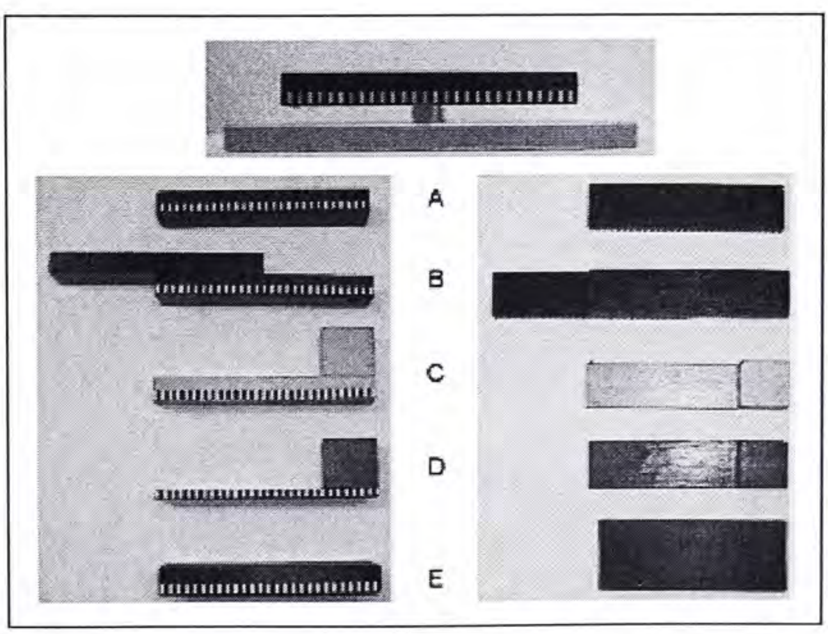

Fig. 2.9 Stimuli used in the Krist et al. (2005, p. 187) experiment

It seems that Krist et al.'s (2005) attempt at cancelling the block-balancing task's empirical support for the RR model is convincing, at first glance; however, it is not the case on closer inspection. Krist et al. argued that the RR model implies that the performance of children must exhibit a U-shaped curve in the course of development. They pushed this statement so far as to say that even response time and frequency of correct adjustments should follow a U-shaped pattern. They had totally mistaken what Karmiloff-Smith meant in her postulation about the U-shaped curve. Karmiloff-Smith (1992, p. 20) said: "The temporary disregard for features of the external environment during phase 2 can lead to new errors and inflexibilities. *This can, but does not necessarily* [italics added], give rise to decrease in successful behaviour—a U-shaped performance curve." Karmiloff-Smith said herself that a U-shaped performance curve is not essential in the RR model, so Krist et al. misunderstood the role of a U-shaped curve in the RR model.

Furthermore, Karmiloff-Smith (1992, p. 88) said the following: "It is also important to note that 6 year olds retain the level-I representations. Ask them to close their eyes and they can balance all the blocks easily." Karmiloff-Smith did not say that 6-year-old children cannot balance asymmetrical blocks. What she said is that the 6-year-old children were not aware of themselves being constrained by the unconscious geometric-centre theory, therefore they failed to balance asymmetrical blocks when this problem was presented to them in a way that required use of conscious inference to solve the problem. The crucial point is how children develop their conception of the task, not whether they can behaviourally perform the task. Therefore, it is important to create an atmosphere that allows children to freely construct their own theory about balancing. If the experimental setting gives children an impression that they should be able to do it, the result would be affected. In the Krist et al. (2005) experiment, part of the instruction is as follows: "Please, try to put each block on this rod so that it will not fall down. I am not going to help you. If you can't do it the first time, that's not a problem. You can try it again if you want" (Krist et al., 2005, p. 187-188). When a child paused after having failed to balance a block, the child was encouraged to try again. If the pause was because the child thought the block could not be balanced, encouragement from the experimenter may have given the participant the impression that the block could be balanced. This setting clearly discouraged children to construct their own theory of balancing and instead they felt they should be able to balance all of the blocks. The experiment became a test for the children's behavioural ability, not exploration of conceptual change. It is not surprising that the U-shaped curve disappears.

Krist et al. criticized Pine and Messer (1999, 2003) for repeatedly asking children for an explanation and they argued as follows:

> "Children of this age tend to elaborate the geometric-center theory or a similar rule if they are induced to think about the question of why some blocks balance and some do not. Note that such a question is actually misleading because it erroneously presupposes that there are blocks which cannot be balanced." (Krist et al., 2005, p. 192)

Therefore, they did not ask for an explanation from the children, and this further discouraged the elicitation of conceptual change in the experiment. Moreover, without verbal explanation provided by the participants, there was no way of knowing the participants' explicit understanding of the task.

Though there were some problems found in the experiment of Krist et al., their empirical finding of "no U-shaped performance curve" cannot be simply dismissed, because the U-shaped performance curve was related to the existence of the geometric-centre theory. It is the geometric-centre theory that causes the performance to drop and the implicit version of the geometric-centre theory is an example of the level-E1 representation. In the current study, Study One is designed to create an atmosphere that allows children to freely construct their own theory about balancing and the issue of whether the U-shaped performance curve exists will be investigated.

2.6 The RR model and the conceptual development of probability

To the best of the author's knowledge, the RR model has not been applied to the study of children's development in probability concept. In this section, I will compare two key predictions of the RR model against the literature in this domain to see whether these predictions match the existing empirical findings.

2.6.1 Piaget's theory of conceptual development of probability

Piaget and his colleagues were the first group of researchers to study developmental analysis of probability judgment and their theories on this topic dominated the field for many years (Reyna & Brainerd, 1994). Piaget and Inhelder (1975) suggested that preoperational children are unable to distinguish what is possible from what is necessary. Preoperational children are prelogical, ruled by systems of intuitive regulations, and have no concept of chance or probability, so they are unable to use frequency information to make probabilistic judgment. Concrete operational children can discriminate events that would necessarily happen and events that are simply possible, because they begin to conceive sets of mixed elements as totalities formed from quantifiable parts. They can make decisions based on frequency, but do not understand the complex relationships between outcomes. In the formal operation stage, two discoveries help children to structure their system of probabilities. The first discovery is the combinatoric system; that is, the system that can construct all possible operations by using a small number of elements. The second is the discovery of proportions. Added together, the two discoveries enable children to have a complete understanding of probability.

Piaget's theory assumed that intuition is primitive and inaccurate knowledge, and that this kind of representation is replaced after children have advanced beyond the preoperational stage, when they acquire computational strategies to solve a probability problem. In contrast, the RR model predicts that the lower level representation is not replaced and coexists with the higher level representation. Intuitive knowledge is very similar to level-I representation, because the internal process of intuition is not accessible to consciousness and people can only be conscious of the product of intuition. Knowledge that is embedded in intuition cannot be used directly in conscious inferences. Piaget also assumed that the developmental direction of the probability concept should be top-down, because explicit understanding of the particular defining features of the probability concept is the prerequisite of further behavioural improvement. Examples of explicit understanding include the ability of concrete operational children to conceive mixed elements as totalities that are formed from quantifiable parts, and the ability of formal operational children to acquire explicit knowledge about the combinatoric system and proportions. The RR model predicts that the opposite developmental direction is also possible, in which the development of explicit understanding can happen after behavioural improvement.

2.6.2 Coexistence of the higher level and lower level representations

In a review of research on conceptual development of probability, Reyna and Brainerd (1994) mentioned that Lovett and Singer (1991) studied the performance of children and adults when solving probability problems. In the task, participants were shown one of two displays. One depicted a pond with an outcrop of rock, in which the precise value of the length of the rock and the area of water could not be known; they could only be estimated perceptually. The other display contained a box with a number of pots of flowers and spiders in which the number of flowers and spiders could be counted. Participants were required to estimate the probability of an insect landing on the rock in the first display, and the probability of an insect landing on a pot of flowers in the second display. Children and adults performed equally well in both situations. Lovett and Singer (1991) found that in problems where counting could be used, the frequency of using counting to solve the problems among different age groups resembled a inverted U-shaped curve, because both young children and adults were less likely to count. Only 25% of the adults chose to compute the numbers and the remaining adults chose to estimate magnitudes perceptually. Therefore, as the RR model predicts, a lower level representation that supports probability estimations is retained even when a higher level representation is available. The lower level representation supports probability estimations without knowing the actual number, or without performing computation. The higher level representation supports precise and conscious computation.

2.6.3 Direction of development: Bottom-up

In the course of studying probability development, two main paradigms are used. Early research using the choice paradigm seems to agree with Piaget's suggestions (Chapman, 1975; Ross & Hoemann, 1975). In the choice paradigm, usually there were two events: event A and event B. Children were asked to decide which event has the greater probability of happening. For example, there are two red marbles and one blue marble in jar A, and six red marbles and two blue marbles in jar B. When concrete operational children are asked to choose which jar they want to draw from if their aim is to get a blue marble, usually they chose jar B because the number of blue marbles is greater.

Acredolo et al. (1989) used the functional measurement developed by Anderson (1980) to study children's probability estimation. Their participants were 7- to 11-year-old children. In Study One, participants were required to provide their estimation using a ruler, at one end of which there was a happy face and at the other end there was a sad face. A sliding marker was attached to the ruler and participants answered by sliding the marker to a position that they believed to be correct. The participants were told that a teacher wanted to offer every child a jelly bean and each child would draw a jelly bean from a well-shaken bag without looking. Participants were to decide the probability of drawing a jelly bean of a particular colour. Analysis of the participants' answers showed that, children from all of the age groups attended to variations in denominator, numerator, and interaction of these two factors. This finding disagrees with research using the choice paradigm, which suggested that children are only sensitive to changes in numerators. In Study Two, using a similar paradigm, Acredolo et al. found that children did not overweight the effect of numerators. Acredolo et al. suggested that "young children possess tacit knowledge of relations between dimensions well ahead of the point at which they are capable of understanding the underlying mathematics" (p. 934). However, the children were not required to explain their answers in this experiment, therefore the children's explicit understanding of the task and probability concept cannot be determined.

The concept of probability is closely related to the concept of proportion. The bottom-up direction for the development of the concept of proportion received support from a line of research on proportional reasoning. Moore, Dixon, and Haines (1991) asked participants to estimate temperature. The relationship between variables and temperature was determined by this equation $T_F = (Q_1T_1 + Q_1T_2) / (Q_1 + Q_2)$, in which T_1 stands for temperature of the first cup of water, Q_1 is the amount of water in the first cup. T_2 is the temperature of water added to the first cup, and Q_2 is the amount added. T_F is the resultant temperature of the water. The participants were divided into two groups. One group received information in quantitative formation, such as "60°," and the other group received information that encouraged intuitive reasoning, such as "hot." Moore et al. (1991) used different dimensions to measure the quality of the participants' strategy and they found that the quality of children's strategies, particularly for younger children, was higher in the intuitive condition than in the computation condition.

From this literature review, it can be seen that both of the predictions based on the RR model, namely the retention of lower representation and the possibility of acquiring implicit knowledge prior to explicit knowledge, agree with the empirical findings of this domain.

2.7 The RR model and the distinction of explicit-implicit memory

The distinction of explicit and implicit memory (Schacter, 1987), or declarative and nondeclarative knowledge (Cohen, 1984; Squire, 1982), is also found in the human memory system. Moreover, concepts possessed by a person must be implemented in the memory system, so that the concepts can be stored for future use. Both Anderson (1982) and Karmiloff-Smith (1992) pointed out that there is an implicit, procedural layer of conceptual representation. This layer should correspond to implicit memory and procedural memory in the memory literature, and if their models are correct, the description of proceduralized knowledge (Anderson, 1982) and level-I (Karmiloff-Smith, 1992) should agree with the findings in the memory literature.

Normal people use both procedural memory and other types of explicit memory when dealing with different types of tasks. To tease apart the memory systems in order to understand the characteristics of implicit memory only, one of the common methods is to study the performance of people with amnesia. Amnesic participants have deficits in verbal long-term memory (Brooks & Baddeley, 1976) and in memory of recent events, so they are severely impaired when performing tasks that need recall or recognition of recently encountered facts or events. They cannot make use of declarative knowledge (Squire & Frambach, 1990) except for knowledge learnt before the onset of amnesia. Therefore, new knowledge that they acquire should be stored in implicit or nondeclarative memory, and should be in the form of procedures or level-I representation. Researchers found that amnesic participants can: acquire motor skills (Brooks & Baddeley, 1976) and perceptual skills such as reading mirror-reversed words (Cohen & Squire, 1980); show the word priming effect (Schacter & Graf, 1986); and show intact classical conditioning (Daum, Channon, & Canavar, 1989). I am going to focus on the acquisition of cognitive skill since this skill is more closely related to conceptual development.

Glisky, Schacter, and Tulving (1986) trained amnesic participants to manipulate information on a computer screen, so that they could write, edit, and execute simple computer programs. The participants learnt the meanings of terms and the association of operation and command words by the vanishing cue method. For example, the computer presented "A sequence of characters enclosed in quotation marks is called a ________." The participant was required to enter "STRING" in response. If the participant did not know the answer, the initial letter of the answer was given as a hint and more hints were given when required. The training program also gave a hint if the participant committed an error in the process of writing a program. The amnesic participants successfully learnt how to operate the computer and how to write a program. The amnesic participants required much more time to learn than normal participants and their concepts were hyperspecific, so the amnesic participants patients had difficulties in answering general or open-ended questions, like "How do you write a program?" Moreover, they failed to answer questions that were composed of words different from the training phase. Amnesic patients also failed in a transfer task. They had learnt how to perform mathematical operation on a computer, and they knew how to write program to output strings, but when they were required to write a computer program for performing mathematical operation, none of the amnesic patients could do. Control subjects who received the same training experienced no difficulties in performing the same task. The difficulties experienced by amnesic patients were exactly the limitation of inflexibility of level-I representation described by Karmiloff-Smith (1992).

Cohen, Eichenbaum, Deacedo, & and Corkein (1985) investigated whether amnesic participants were able to solve the Tower of Hanoi puzzle, which is shown in Figure 2.10.

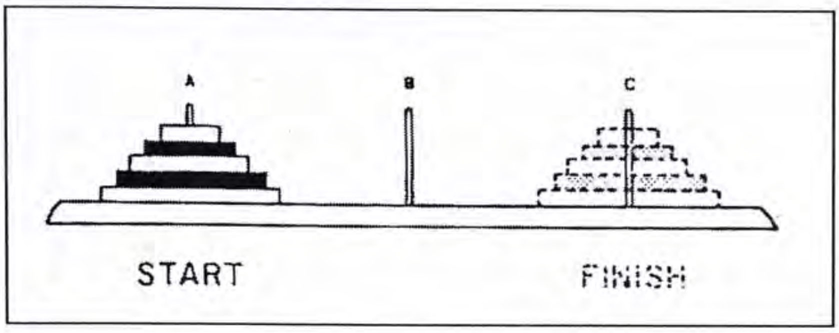

Fig. 2.10 The Tower of Hanoi puzzle. All five blocks are arranged from small to large on peg A in the initial state. All five blocks should be moved to peg C in the goal state. Only one block is allowed to move each time, and a larger block cannot be placed on top of a smaller block. (From Cohen et al., 1985, p. 57)

It was found that the amnesic patients demonstrated normal learning of the puzzle. With more practice, they committed fewer errors and required fewer steps to recover from errors. By the end of the 16-day training, the amnesic patients were able to solve the puzzle from any intermediate step, and the mean number of steps was significantly and substantially reduced. However, this acquisition of skill was hardly revealed by their verbal reports. For instance, in the final test, one amnesic participant H.M. asked the experimenter 13 times whether he was permitted to move a block across two pegs, rather than only one. He claimed to be "stuck" and was unable to think of any legal moves eight times, despite the fact that he was reaching the optimal solution consistently. After solving the problem in the last session, H.M. was asked to provide a verbal account of how to solve the problem. H.M.'s verbal answer involved illegal moves, and when these moves were pointed out by the experimenter, H.M. exclaimed that he could not figure it out, but in fact he had just solved the puzzle by moving the blocks. Improvement of skills was also poorly reflected by the verbal reports of normal participants, because their verbal reports of changes in strategy often occurred much later than the initial occurrence of the behavioural change. This reflects the dissociation between explicit and implicit representation. The inaccuracy in the verbal reports of change in strategy suggests that knowledge can be bracketed by implicit representation, and representation redescription is needed to redescribe embedded knowledge into explicit representation.

Squire and Frambach (1990) studied how amnesic and normal participants behaved in a sugar production task. In the task, participants were required to enter the number of workers to be hired in a sugar factory, in order to achieve a particular sugar production level. The computer generated sugar production levels according to an equation that includes variables of "worker number" and "previous sugar production." It was found that the amnesic patients improved at a normal rate in the first session consisting of 90 trials. After one month, participants came back for session 2, which also consisted of 90 trials. In session 2, the control participants performed much better than amnesic participants. All of the participants were required to complete questionnaires after the task. Neither the control participants nor the amnesic participants were able to answer specific questions about how the task worked and they were quite poor at predicting the computer's response to a particular input. However, the control participants acquired some knowledge about the general strategy for performing the task. This outcome agreed with Karmiloff-Smith's (1992) suggestion that results of redescription are reduced descriptions that lose many of the details, and that practice can lead to redescription of representation from one form to another.

CHAPTER THREE: METHOD

3.1 Study One: The block-balancing experiment

The original block-balancing experiment of Karmiloff-Smith and Inhelder (1974/75) provides an illustrative example for different levels of representational explicitness. It was found that 4- to 5-year-old children possessed level-I representation, 6-year-old children possessed level-E1 representation, and 8- to 9-year-old children possessed level-E3 representation. However, the age difference of representation was challenged by Pine and Messer (1999, 2003) because they found that 6- to 7-year-old children could verbalize geometric-centre theory, which conflicts with the level-E1 representation found in Karmiloff-Smith's study (1974/75). Krist et al. (2005) challenged the findings of previous studies by saying that 6-year-old children's performance was not worse than 4- to 5-year-old children, and no U-shaped performance curve was found in their study.

One crucial question involved in this experiment is whether the geometric-centre theory exists in children's minds, and in what form it exists. If an implicit form of geometric-centre theory exists, it can be observed from a child's consistent pattern of behaviour. Although children should possess the ability to balance asymmetrical blocks, their performance falls short of their actual competence.

Since children are not aware of such implicit theory, they attribute their failure to features of the blocks, rather than to the placement position. Children must be consciously aware of this theory in order to falsify its applicability for the asymmetrical blocks, then their performance is improved. Thus, children may demonstrate a U-shaped performance curve across the span of their development. In this study, the form of geometric-centre theory held by children was investigated.

Geometric-centre theory is an intermediate concept that children possess, before they develop a naïve version of the law of torque, which explains the balance point in relation to weight and length. Geometric-centre theory is one concept that can be used for balancing symmetrical blocks. The naïve version of the law of torque is another more generalized concept that explains how both symmetrical and asymmetrical blocks can be balanced. Though these two concepts are related and have an overlapping area of application, they are different entities in the mind of children; therefore, the two concepts go through developmental phases independently. It is possible for the intermediate concept to be represented explicitly, while the other more generalized concept is still being represented implicitly. It is also possible for one concept to develop in the top-down direction, while the other concept develops in the bottom-up direction. The developmental pattern of these two concepts will be investigated in this study.

3.1.1 Research Questions

1. Can the difference between implicit and explicit representation be observed in the task of block balancing? What are the characteristics of these representations?
    1.1) Is there a stage during which children can balance symmetrical blocks and asymmetrical blocks, but are unable to provide correct verbal explanation?
    1.2) In what form does an implicit geometric-centre theory exist?
    1.3) Will the geometric-centre theory be mentioned by children spontaneously in their verbal explanation of their failure or success in the balancing task?
    1.4) When the experimenter explicitly asks the children whether they are aware of their initial placement pattern, can they give a correct answer?
    1.5) If children are aware of the geometric-centre theory, are they aware of the relation between this implicit theory and the success or failure of balancing?

2. Does bottom-up learning occur in the experiment?
    2.1) Does implicit representation occur before explicit representation among children?
    2.2) How do children's explicit explanations change across trials?
3. Is there a U-shaped performance curve in children's block-balancing attempts?
    3.1) Will children be able to balance asymmetrical blocks in earlier trials, but fail to replicate this success in later trials?
    3.2) Among the three age groups of 4 to 5 year olds, 6 to 7 year olds, and 8 to 9 year olds, will the middle group perform the worst?
4. Both in terms of performance and conceptual understanding, what is the developmental pattern revealed in this block-balancing task?
    4.1) How does an individual develop across trials?
    4.2) How does each age group differ from the others?

3.1.2 Participants

All participants were recruited from a co-educational school located in Ho Man Tin, Kowloon. The school had both primary and kindergarten sections, and most of the students have a middle-class background. Three age groups were included in this study: 4 to 5 year olds, 6 to 7 year olds, and 8 to 9 year olds. Participants of the 4- to

5-year-old age group all came from K3 of the kindergarten section, and their mean age was 5 years 9 months. Participants of the 6- to 7-year-old age group all came from primary 1 of the primary section, and their mean age was 7 years 1 month. Participants of the 8- to 9-year-old age group all came from primary 3 of the primary section, and their mean age was 9 years 1 month. This age range was set after considering Karmiloff-Smith and Inhelder's (1974/75) experiment. According to their experiment, there were performance differences between these three age groups. There were eight participants in each age group, half of them male and the other half female. A total of 24 participants were tested. The same group of children participated in both Study One and Study Two. Half of the participants in each age group attempted Study One first, and the other half attempted Study Two first. Consent from parents and children was obtained before the commencement of the experiment.

3.1.3 Materials

As shown in Figure 3.1, an A4 size (30 cm x 22 cm) wooden board was used as the stage for performing the block-balancing task. A 30 cm x 0.5 cm x 1.5 cm narrow wooden rod was fixed to the wooden board to act as the fulcrum. Eight wooden blocks served as stimuli. The four blocks that balanced at the geometric centre, were referred to as symmetrical blocks, and the other four blocks that balanced off centre, were referred to as asymmetrical blocks.

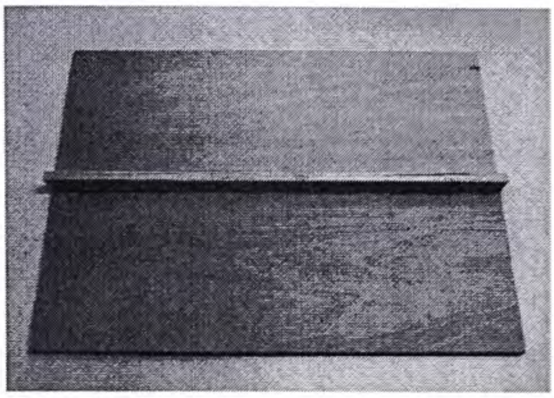

Fig. 3.1 The plate for the block-balancing task

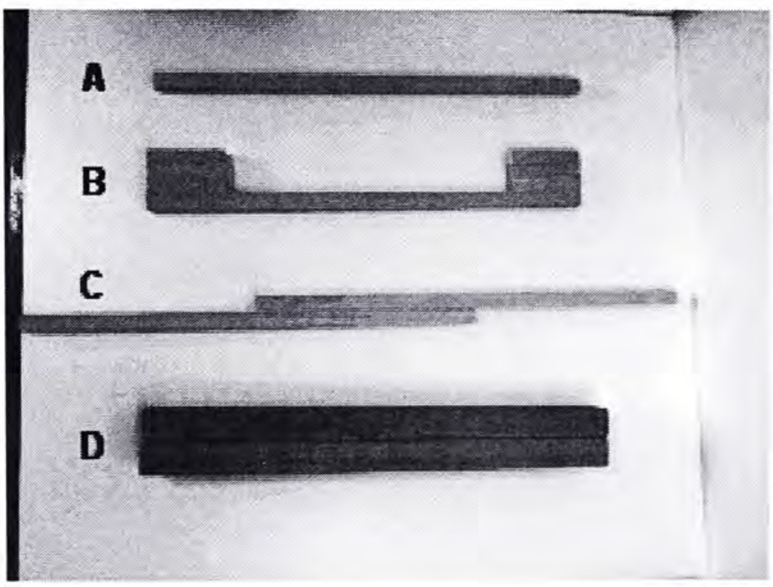

Fig. 3.2 The symmetrical blocks

Blocks A, B, C, and D were symmetrical blocks (see Figure 3.2). Block A measured 30 cm x 1 cm x 1 cm and had nothing attached to it. Block B was made by attaching three small bricks of the same size (5 cm x 1 cm x 1 cm) at both end of a block that was identical to block A. Block C was made by sticking together two blocks that were identical to block A, with an overlapping area of 15 cm. Block D was made by sticking together two wooden blocks (30 cm x 3.5 cm x 2 cm) so that the two blocks overlapped completely.

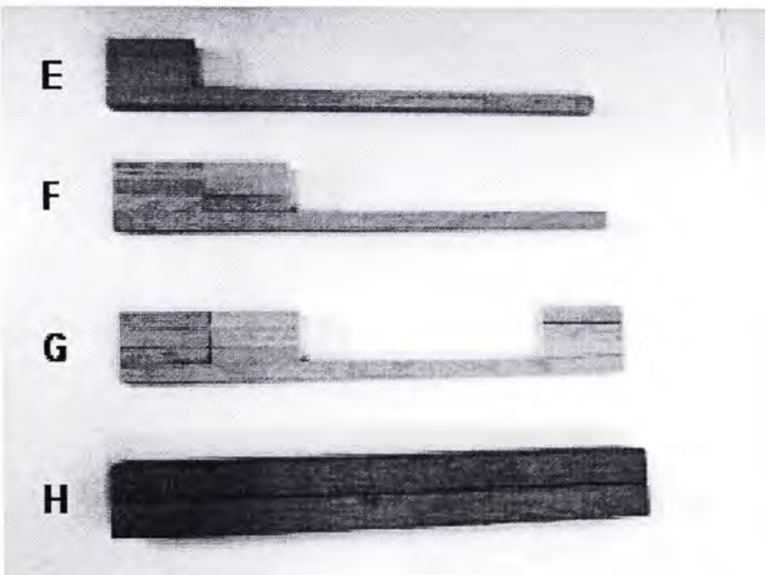

Fig. 3.3 The asymmetrical blocks

Blocks E, F, G, and H were asymmetrical (Figure 3.3). Block E was made by attaching three bricks on only one end of Block A. Block F was made by attaching six bricks on one end of Block A. Block G was made by attaching three bricks on one end of Block A, and six bricks at the other end. Block H was made by sticking together two wooden blocks (30 cm x 3.5 cm x 2 cm) so that the two blocks overlapped completely, but a metal weight was placed between the two wooden blocks, so that one side was heavier than the other side although both sides appeared to be the same.

Blocks A, C, E, and H were adopted from Karmiloff-Smith and Inhelder's (1974/75) experiment. Block B, F, and G were adopted from Messer et al.'s (1998) study. Since three of the asymmetrical blocks carried additional bricks, these bricks could be a feature that attracts the children's attention. Block B could act as a potential counter example for children who think that the existence of additional "bricks" is the reason why the block cannot be balanced. Block D and H look identical, the width and height of these two blocks is larger than other blocks. Block D could act as a potential counter example for children who think that the dimension of the block is the reason why the block cannot be balanced.

Type G in Karmiloff-Smith and Inhelder's (1974/75) experiment was not included, because it cannot be balanced without the use of a counter weight. Also, none of the block-balancing experiments that were conducted after Karmiloff-Smith and Inhelder's experiment used this type of block (Krist et al., 2005; Messer et al., 1998; Pine & Messer, 1999, 2003).

3.1.4 Design and procedure

The experiment was divided into four parts: Phase One Test, a free play period, the Prediction Task, and the Phase Two Test. The whole study lasted around

30-35 minutes. Participants performed the block-balancing task in the Phase One Test and the Phase Two Test, using the same set of blocks and presentation order. The whole process was recorded by digital video camera.

At the beginning of the Phase One Test, the participant was given the following instruction: "Today we are going to play with some wooden blocks. I am going to give you some blocks, I want you to try balancing them over here (point to the wooden board), so that they will not fall. Please don't put it vertically (with demonstration). You need to balance it horizontally. It may be the case that you can balance all of the blocks or it is also possible that some of the blocks cannot be balanced. If you think that a particular block cannot be balanced, you can tell me about that. This game is very challenging, never mind if you find it difficult. Just try to do your best."

Incorrect vertical placement was demonstrated, but horizontal placement was not demonstrated to avoid any influence on the participants' initial placement point.

Each stimulus block was attempted once in the Phase One Test. Only the current test block was visible to the children to avoid distraction. The order for presenting the blocks was A, E, B, D, F, G, C, and H. Earlier blocks were usually easier to balance than later blocks, so participants would not become frustrated and lose motivation for continuing the experiment.

The initial placement position of each block was recorded. Participants were allowed to try as many times as they liked. If they chose to give up or declare that the block could not be balanced, the experimenter asked, "Why didn't the block balance just now?" and the explanation was recorded for analysis.

After all of the eight blocks had been attempted once, the Phase One Test ended and the free play period of 5 minutes began. Participants were given the following instruction: "Good, you have tried very hard in the test. Now, I will give you all of the blocks for you to play with for five minutes. Later I will test you again in the way we did just now. You can try again to balance the blocks that you did not succeed in balancing just now." In the free play period, all of the blocks were given to the participants. After the instructions had been given at the beginning of the free play period, the experimenter did not intervene with participant's activities any more. There was also a free play period in Pine and Messer's (2003) experiment. The aim of such a free play period is to allow children to freely explore and to construct their theory about balancing. As mentioned before, the aim of this experiment is not to assess children's behavioural ability to balance the wooden blocks, but to study how their concept of this task developed.

The Prediction Task started after the 5 minutes' free play. The experimenter collected all of the blocks and put them into one group. The experimenter then took out two cards, one printed with a big tick and one printed with a big cross. The card printed with a tick was placed on the right, the cardboard with a cross was placed on the left. The participant was given the following instruction: "I want you to classify this pile of blocks. If you think that a certain block can be balanced, place it over the card with a tick. If you think that a certain block cannot be balanced, place it over the card with a cross." After the participant finished the sorting process, the experimenter recorded the results.

After the Prediction Task, the Phase Two Test began. The experimenter collected all of the blocks and reran the test using the same procedure as in the Phase One Test. At the end of the Phase Two Test, the experimenter asked three more questions:

1. Is there a position on the block that you usually place over the fulcrum?

2. Did you notice your habit before I asked you about it?
3. Do you think the way you usually place the block will affect whether a block can be balanced?

After answering all of these questions, the child was thanked for his or her participation and a souvenir was given to each participant.

3.1.5 Analysis

In some studies of conceptual developmental, it is taken for granted that behaviour and verbal reports are just two sides of the same coin, both of them reflect the concept(s) possessed by the children. From the perspective of the RR model, the representation that is responsible for behaviour and the representation that is responsible for explicit knowledge may both represent the content of the same concept, but they may belong to different levels of representation. Different representations may develop at different paces and may give rise to a difference in behavioural and verbal performance. To see whether this kind of difference in performance really existed, the behavioural performance and performance of explicit knowledge were first analyzed independently. After building up some basic dimensions for measuring both types of performance, these two kinds of performance were compared to see if there was any difference between them.

3.1.5.1 Behavioural performance

According to the RR model, only knowledge that is represented at the E2/3 level can be verbalized. Knowledge represented at levels lower than E2/3 cannot be verbalized and can only be observed from behavioural performance. In this study, behavioural performance of the block-balancing task was measured in three ways: Success Score, Initial Middle Placement Score, and Geometric Centre Area Placement Score. The first two ways of measurement are very common in the block-balancing paradigm, for example, as can be found in the work of Messer et al. (1998) and Krist et al. (2005). To the best of the author's knowledge, the third way of measurement cannot be found in the current block-balancing paradigm literature.

The Success Score was calculated by counting the number of successfully balanced blocks. The Initial Middle Placement Score was calculated by counting the number of times the block was initially placed at the geometric centre. Both scores were analyzed using analysis of variance (ANOVA).

The Geometric Centre Area Placement Scores (GCP Scores) of symmetrical and asymmetrical blocks were obtained by counting the number of blocks placed in the geometric-centre area only throughout the trial for each block. It should be noted that "geometric-centre area" does not only mean the exact point of geometric centre, but also includes an area close to the geometric centre. The reason for choosing "central area" rather than the centre point is that even if young children have the intention of placing the block at the centre, the placement position may not be the exact geometric centre because of their immature motor ability. ANOVA of GCP Scores were also carried out.

If geometric-centre theory really exists, it leads to failure in balancing asymmetrical blocks by constraining block placement positions throughout the trial, not only the initial placement. Therefore, performance in the rest of the trial should also be considered. Children who possess this theory try to balance all of the blocks in the geometric-centre area only, but the balance point for asymmetrical blocks (except block G) are not in this area, so that is why these children cannot balance asymmetrical blocks. In other words, persistent central placement is the direct effect of geometric-centre theory, and success or failure in block balancing is the result of persistent centre placement. Success in block balancing can be affected by factors other than geometric-centre theory. For example, younger participants tend to make less subtle changes in placement. Sometimes the placement positions they tried were roughly correct, and they failed only because the placement was not precise enough. Therefore, the GCP Score should be a more sensitive measure of geometric-centre theory compared to Success Score or Initial Middle Placement Score.

If a block is only placed in the central area throughout the trial, 1 point is added to the GCP Score for symmetrical or asymmetrical blocks. To maintain consistency between different behavioural performance measures, the following rules are defined:

1. If a block does not score a point for the initial centre placement score (that means the initial placement position of that block is off centre), the GCP Score of the block is zero.
2. Except for Block G, if an asymmetrical block is successfully balanced, its GCP Score is zero.

Block G is exempted from the second rule because its balance point is within the central area, though it is not at the geometric-centre point.

One of the research questions was to investigate whether the implicit geometric-centre theory exists and whether younger children persistently try to balance all of the blocks in the geometric-centre area. In this study, the behavioural pattern of each participant should be considered individually. Test Phase One or Two is classified as exhibiting a behavioural pattern of geometric-centre theory (BGeo) if the GCP Score for an asymmetrical block in that test phase is greater than or is equal to 3.

The geometric-centre theory and the naïve version of the law of torque are two different conceptual entities that may go through different developmental paths. Therefore, the behavioural pattern reflecting the naïve version of the law of torque (BTorque) should also be analyzed. The Phase One or Phase Two test is classified as exhibiting BTorque if 3 or more of the trials for asymmetrical blocks fulfil both of the following two requirements:

1. The block is successfully balanced.
2. The initial placement position is on the correct side of the block, that is, the initial placement position is on the heavier side of the block.

It should be noted that BGeo and BTorque are mutually exclusive for a priori reasons. BGeo requires three or more central-area-only placements, which guarantees at least two failures in balancing asymmetrical blocks (asymmetrical Block G balances within the central area, therefore the minimum number of failures is not three but two). BTorque requires participants to successfully balance at least three of the four asymmetrical blocks in a test phase. Therefore, it is impossible for a test phase to exhibit both BGeo and BTorque at the same time.

3.1.5.2 Performance that demonstrates explicit understanding

According to the RR model, knowledge that can be verbalized is represented at the E2/3 level. In this experiment, children's explicit understanding was studied in two ways by examining their performance in the Prediction Task and their verbal explanations provided in Phase One and Phase Two tests.

In the Prediction Task, the participants' predictions of whether the blocks can be balanced was analyzed. This task was designed to enquire the accessibility of the concept(s) involved in the block-balancing task. If a certain concept is represented at level-I, then participants should not be able to use this concept for making a prediction, because questioning and balancing involve different modalities, and level-I representation does not allow interdomain or intradomain links to be established. If participants made use of a certain concept in making a prediction, then some kind of cross-modality accessibility was shown and the representation should be at a level above that of level-I. If a participant consistently predicts asymmetric blocks as impossible to balance, and the explanation provided by the participant does not mention the "middle point," it can be interpreted that the geometric-centre theory is an implicit assumption that the participant is not aware of.

The verbal explanations obtained in test Phases One and Two were classified into different categories, which helped to reveal the pattern of development, and this classification was also the foundation for further analysis. During the pilot study reported in section 3.4, five types of verbal explanations were observed:

1. No explanation.
2. Physical properties: An explanation that attributed success or failure to physical properties of the block, like weight or size.
3. Geometric-centre theory: An explanation that included "middle point" or "length of both sides is equal."
4. The naïve version of the law of torque: An explanation of the relationship between weight and balance point.
5. Other explanations (Idiosyncratic).

After examining the empirical data of the current study, the explanation type that was related to the physical properties of blocks were further divided into 3 subtypes: P1, P2, and P3, in order to reflect the variations of the children's explanations more accurately.

There are eight categories in the new system. To assess the interrater reliability, transcripts and video recordings of 10 participants (41.7% of all trials) were given to a second rater to independently perform classification. The first rater was the researcher. The second rater was a PhD student. The interrater reliability was evaluated using Cohen's Kappa. With $\kappa = .887$, the degree of agreement was almost perfect (strength of agreement based on Landis and Koch, 1977)).

Here are the classification requirements of the eight categories:

1. ***No explanation (N)***: Participants said that they did not know.
2. ***Physical properties (P1, P2 or P3)***: Explained by the physical properties of the block. Explanations that belong to this category were further classified into one of the following three subtypes:
   i. P1: Explained by the size or weight of the whole block.
   ii. P2: Explained by the equality or inequality of size or weight of the two

sides of the block (divided by the geometric centre).

iii. P3: Explanation did not belong to the above two subtypes.

3. ***Geometric-centre theory (G):*** Explained the success of balancing a symmetrical block by mentioning that the block could be balanced at the geometric centre. Explained the failure in balancing an asymmetrical block by mentioning that the block could not be balanced at the geometric centre. The geometric centre or expressions that have similar meaning must be explicitly mentioned in the explanation.

4. ***Balance point related (B)***: Explanation fulfils only one of the following conditions:

   i. Explained that a balance can be achieved if two sides of the block (divided by the balance point, not by the geometric centre) carry the same weight, but the concept of balance point was not explicitly mentioned.

   ii. Used the concept of a balance point in the explanation.

5. ***The naïve version of the law of torque (T)***: Mentioned either the complementary relationship between the length of one side of the block and the weight of the other side of the block, or explained by mentioning weight, balance point, and how the relationship of the two contributes to balancing.

6. ***Others – Idiosyncratic (O)***: Explanation did not belong to any of the above categories.

To have a balanced picture of the distribution of the explanation types, the distribution was analyzed in two ways: the frequency of occurrence among the trials in each age group, and the number of participants who had used a certain type of explanation in each age group. If a certain explanation was found to have a low frequency of occurrence among the trials, this may be because all of the participants had tried the explanation, but each of them used it in very few trials. Alternatively, it may be because the explanation was used by a few participants consistently throughout the experiment. Therefore, if only the frequency was considered, the distribution may be misunderstood.

The conceptual understanding reflected by each explanation type was analyzed in detail. Explanation type T (the naïve version of the law of torque) was the most correct type of explanation in the classification system, so what kinds of explanations preceded and were followed by the discovery of the type T explanation were explored.

3.1.5.3 Comparison of behavioural performance and performance that demonstrates explicit understanding

Two major issues were investigated by analyzing the data of both behavioural performance and performance based on explicit understanding: geometric-centre theory and the direction of development.

Karmiloff-Smith (1992) believed that 6-year-old children implicitly possessed the geometric-centre theory, which led to their persistent failure in balancing asymmetrical blocks. This implicit geometric-centre theory was used to prove the existence of the E1 level. Therefore, the issue of whether implicit geometric-centre theory exists has theoretical importance to the RR model.

If a participant had an implicit form of geometric-centre theory, this could be observed by the participant's behavioural pattern, which could be measured by BGeo in this study. If a participant had explicit knowledge about the geometric-centre theory, this knowledge should be found in the participant's explanation within a test phase and the explanation was categorized as a type G (Geometric-centre theory) explanation. Also, at the end of the experiment, the experimenter explicitly asked the participants whether they were aware of their habit of initial placement, and if the placement position affected whether a block could be balanced or not.

To study the developmental changes that occurred within the experiment over time, four periods of time were defined, and in this study are referred to as "time points." T1 was test Phase One, T2 was the Prediction Task, T3 was test Phase Two and T4 was the last part of the experiment in which the experimenter explicitly asked questions about the initial placement position and its relationship to block balancing. For both T1 and T3, four kinds of possible situation may exist, therefore four kinds of label were used:

1. **Implicit geometric-centre theory (IG)**: If BGeo was found, but no explanation type G was found within the eight trials of T1 or T2, this time point was interpreted as reflecting the existence of implicit geometric-centre theory, and labelled as "IG".
2. **Explicit geometric-centre theory (EG)**: If BGeo and explanation type G were both found, the participant possessed explicit geometric-centre theory at that time point, and label "EG" was used.
3. **Explicit geometric-centre theory only (E Only)**: If BGeo was not found but the participants tried to use explanation type G to explain balancing an asymmetrical block for at least one time (but not symmetrical blocks, since symmetrical blocks

do balance at the centre), the time point was labelled as "E Only."

4. **No geometric-centre theory found (N)**: If BGeo was not found and the participants did not try to use explanation G, that meant the participant did not exhibit any sign of geometric-centre theory, and the time point was labelled as "N."

For T2, that is, the Prediction Task, two labels were used:

1. "≥3": If the participant predicted that three or more asymmetrical blocks could not be balanced.
2. "-": If the above condition was not met.

Because the T4 time point follows immediately after T3, the performance of T3 affects the labelling categories for T4.

1. **IG**: If T3 is IG, and T4 fulfils one of the following conditions:
   i. The participant's answer reflected that the participant was not aware of his/her persistent initial middle placement, ***or***
   ii. The participant was aware of the placement habit, but do not think that it related to the experience that the asymmetrical blocks could not be balanced.

2. **EG**: If T3 is EG, or T3 exhibits BGeo, and T4 fulfils both of the following conditions:

    i. The participant's answer reflected awareness of the persistent initial middle placement, ***and***

    ii. The participant believed that the asymmetrical blocks could not be balanced because those blocks fell when being placed at the centre.

3. **E Only**: If T3 is E Only, or T4 fulfils both of the following conditions:

    i. The participants' answer reflected awareness of the persistent initial middle placement, ***and***

    ii. The participant believed that the asymmetrical blocks could not be balanced because those blocks fell when being placed at the centre.

4. **N**: If T3 is N, and the participant's answer does not reflect the belief that asymmetrical blocks cannot be balanced because those blocks fall when placed at the centre.

One of the major objectives of the current study is to use the explicit-implicit dimension for studying development. How development happens across different levels of knowledge representation will be investigated. Viewing from this explicit-implicit dimension, development can be divided into two main directions:

top down and bottom up. The top-down direction of development refers to the process that knowledge in an explicit level guides the development of an implicit level. The bottom-up direction of development involves turning knowledge embedded in an implicit level into explicit conceptual knowledge. If top-down development takes place, explicit knowledge is a precondition of behavioural success. For example, in the participants studied by Anderson (1982), if the participants had no explicit knowledge of the postulates, they could not solve the geometric problems by applying these postulates. If explicit knowledge is the precondition of behavioural success, then without explicit knowledge, no success is possible. That means, the conditional probability of P (Behavioural success | Explicit knowledge does not exist) and P (Explicit knowledge does not exist | Behavioural success) should both be 0. The above two probabilities were calculated to investigate the relationship between success in balancing individual blocks and giving a correct explanation, as well as the relationship between BTorque and a naïve version of the law of torque.

3.2 Study Two: The probability-estimation experiment

Previous researchers have found that people can perform probability estimation intuitively without explicit computation (Lovett & Singer, 1991), and young children can have tacit knowledge about probability (Acredolo et al., 1989). This intuition or tacit knowledge is very similar to knowledge represented at level-I of the RR model, but this requires verification because past studies only focused on participants' behavioural performance. It should be noted that the ability to perform a task and the participants' conception of a task are two different issues. Therefore, in this study, both the results of probability estimation and verbal explanations were studied; the results were viewed as two sources of data, each of which has equal importance for understanding the characteristics of the children's minds.

The RR model predicts that the probability concept can be acquired from the bottom-up direction, and might aid in describing how the implicit representation of this concept can be transformed to an explicit representation. These developmental patterns must be captured when development is actually taking place, therefore trial-by-trial analysis was used in this study.

3.2.1 Research Questions

1. Can the distinction of implicit and explicit representation be observed in the task of probability estimation? What are the characteristics of these representations?

    1.1) Does the behavioural performance of probability estimation develop at the same pace as the performance that involves explicit understanding?

1.2) Are any thoughts related to a certain concept of probability mentioned spontaneously by the children in their estimation task?

1.3) In response to the experimenter's request for an explanation for the probability estimation in each trial, is there any explicit representation of a probability concept?

1.4) In response to the experimenter's counter argument and probing questions, what are the characteristics of the children's explicit representation of the probability concept?

2. Does bottom-up learning happen in the experiment?

2.1) Does implicit representation occur before explicit representation?

2.2) How do children's explicit explanations change across the trials?

3. Is there a U-shaped performance curve in children's performance of the probability estimation?

3.1) Among the three age groups of 4 to 5 year olds, 6 to 7 year olds, and 8 to 9 year olds, do the middle group perform the worst?

4. Both in terms of performance and conceptual understanding, what is the developmental pattern revealed in this probability estimation task?

4.1) How does an individual develop across the trials?

4.2) How does each age group differ from others?

3.2.2 Participants

All of the participants were recruited from a co-educational school located in Ho Man Tin, Kowloon. The school has both primary and kindergarten sections and most of the students have a middle-class background. Three age groups were included in this study: 4 to 5 year olds, 6 to 7 year olds, and 8 to 9 year olds. Participants of the 4- to 5-year-old age group all came from K3 of the kindergarten section and their mean age was 5 years 9 months. Participants of the 6- to 7-year-old age group all came from primary 1 of the primary section and their mean age was 7 years 1 month. Participants of the 8- to 9-year-old age group all came from primary 3 of the primary section and their mean age was 9 years 1 month. The youngest group of participants in Acredolo, O'Connor, Banks, and Horobin (1989) was 7 year olds. The 4- to 5-year-old group was added to see if they also possess implicit representation of probability. In the pilot test (see section 3.4), 4- to 5-year-old participants demonstrated that they were capable of understanding and tackling the task and no floor effect was observed. There were eight participants in each age group, half of them male and the other half female. A total of 24 participants were tested. The same group of children participated in both Study One and Study Two. Half of the participants in each age group attempted Study One first, the other half attempted Study Two first. Consent from parents and children was obtained before the commencement of the experiment.

3.2.3 Materials

The probability estimation task itself can be quite boring, so children may lose motivation and refuse to concentrate on the task. Therefore, the experimental instruction was presented in the form of a story. Three hand puppets were used to attract children's attention. In the pilot study, many participants showed excitement when I took out a new hand puppet. At the beginning of the pretest, a sheep hand puppet was shown and children were told the following story. Little Sheep lived in the Sun Kingdom. All of the animals in the Sun Kingdom ate grass. There were two types of grass in this kingdom, one type was delicious and green in colour, the other type was bitter and brown in colour. Little Sheep went to a party one day, but there was not enough delicious grass. So the host of the party decided to distribute grasses by asking the animals to draw grass from a bag. Each animal had to close its eyes and draw grass from a transparent plastic bag. The children were asked to estimate how happy Little Sheep would be when she saw different combinations of delicious grass and bitter grass in the bag she would draw from. At the beginning of the training phase, a hand puppet of a horse was shown, and the children were told that

Little Horse also loved to eat green grass and hated brown grass. It was now Little Horse's turn to draw grass from the bag and the children were required to estimate Little Horse's happiness. This time Little Horse told the children the correct answer. At the beginning of the posttest, a hand puppet of a pig was shown. The children were told that Little Pig wanted to learn to perform estimations, and the participant was asked to teach Little Pig if she asked questions.

The participants indicated their probability estimations by moving an arrow that was attached to a plastic Answer Board (see Figure 3.1). Each of the 10 marks on the Answer Board were marked with a number. The first mark on the left was marked with "0," and an unhappy face was placed above the 0, the following marks were "1, 2, 3, 4, 5, 6, 7, 8," and "9." The last mark was "10," and a happy face was placed above the 10. Children indicated their answers by moving the arrow to the number they preferred.

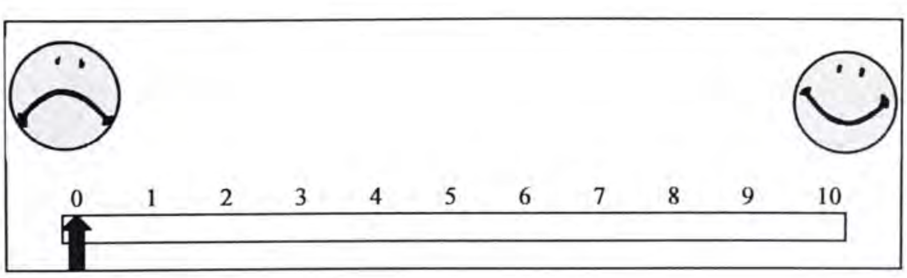

Fig. 3.4 The Answer Board for indicating an answer

Stimuli were presented by a notebook computer using Microsoft PowerPoint. Each question was presented in one slide. Green-coloured grass images represented delicious grass and brown-coloured grass images represented bitter grass (see Figure 3.2).

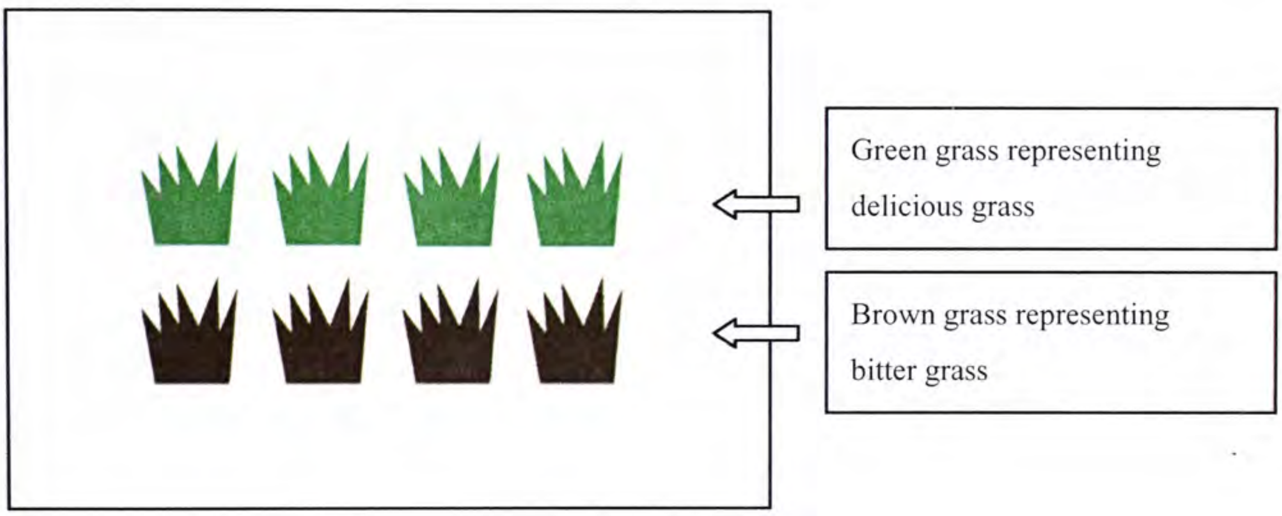

Fig. 3.5 An example of the stimuli slide

Both the pretest and the posttest contained the five questions that are shown in Table 3.1. In both the pretest and the posttest, the first two questions had a different numerator (number of delicious grass images) and a different denominator (total number of grass images), but the probability value of the questions was the same. For example, in the pretest, question one was 1/2, question two was 4/8, so the probability value is the same. It was found in Acredolo et al.'s (1989) study that children often failed to understand that the true probability is the same in such a variation of numerator and denominator. The third and fourth questions in the pretest and the post test were another pair of similar questions; in both pairs, the one with the small numerator had higher value. This was included to test the claim of some researchers that concrete operational children cannot handle this kind of question (e.g., Chapman, 1975; Ross & Hoemann, 1975).

Q no	Pretest	Value	Posttest	Value	Remarks
1	1/2	5	5/10	5	Q1 and Q2 are the same in probability
2	4/8	5	3/6	5	value, but the numerator and denominator are different
3	2/3	6.7	5/9	5.6	For Q3 and Q4, the one with the smaller
4	3/9	3.3	3/4	7.5	numerator has a higher probability value
5	1/5	2	8/9	8.9	-

Table 3.1 Questions for the pretest and the posttest

The questions used in the training phase are shown in Table 3.2. The first three questions have the same numerator, but of different denominators. The third and fourth questions share the same probability value, but their numerators and denominators are different. The fourth and fifth questions, as well as the sixth and seventh questions, are two pairs that share the same denominators but with different numerators, so their probability values are different.

Q no	Training	Value	Remarks
1	4/5	8	Q1, Q2, and Q3 share the same numerator
2	4/7	5.7	
3	4/10	4	
4	2/5	4	Q3 and Q4 share the same value, but their numerators and denominators are different
5	2/10	2	Q4 and Q5 share the same numerator
6	6/10	6	Q6 and Q7 share the same numerator
7	6/7	8.6	
8	7/10	7	

Table 3.2 Questions for the training phase

The three hand puppets were used in the experiments were a sheep, a horse, and a pig.

3.2.4 Design and procedure

The experiment was composed of three phases. The pretest contained five questions, the training phase contained eight questions, and the posttest contained five questions. The whole study lasted around 30-35 minutes, and the process was recorded by digital video camera. The aim of the pretest was to test the participants' initial concept of probability. In the training phase, feedback was provided to the children, and the questions were grouped together in terms of their characteristics; it was hoped that this would help the children to redescribe implicit knowledge about probability into explicit knowledge. In the posttest phase, counter suggestions to the participants' answers were raised to elicit more verbal responses and to test whether the participants' beliefs were consistent.

3.2.4.1 The pretest

At the beginning of the experiment, the experimenter gave the following instruction to the children.

> Little Sheep lives in the Sun Kingdom. All of the animals living in the Sun Kingdom eat grass. In the Sun Kingdom, there are two types of grasses. One type of grass is bitter and it is brown in colour (demonstrated with the PowerPoint slide). The other type is very delicious and it is green in colour (demonstrated with the PowerPoint slide). Today, Little Sheep goes to a party, but there is not enough delicious grass there. To avoid animals fighting for the delicious grass, the host of the party has decided to put the grasses into a transparent plastic bag (the experimenter took out a transparent plastic bag containing cards printed with green grass or brown grass), and each animal has to draw grasses from the plastic bag. Before the animal draws from the bag, it has to close its eyes (the experimenter covered her eyes with her hands), the bag would be well shaken (the experimenter shook the bag), and the animal

has to draw with its eyes closed (the experimenter drew a card from the bag). I would like you to guess how happy Little Sheep will be when it sees different combinations of numbers of delicious grass and bitter grass drawn from the bag.

The experimenter then demonstrated how to use the Answer Board.

> If Little Sheep finds that the grasses in a certain bag is like this (the experimenter showed a slide with all bitter grasses), then she will feel very unhappy (the experimenter moved the arrow on the Answer Board to 0). If Little Sheep finds that the grass in a certain bag is like this (the experimenter showed a slide with all delicious grasses), she will be very happy (the experimenter moved the arrow on the Answer Board to 10).

In the experiment, some of the participants spontaneously moved the arrow to 10 before the experimenter tried to do so. The experimenter then asked whether the participant had any questions. If there were no questions, the experiment would then start.

Every time a new question was shown on the slide, the participant was asked "How many bitter grasses are there?" and "How many delicious grasses are there?"

The aim of asking these questions was to prevent misrecognition, because the youngest participant in the pilot test wrongly recognized the grass types in two trials. He mistakenly recognized brown grass as green grass. After confirming that the images have been correctly recognised, the experimenter said, "Little Sheep is going to draw from this bag. How happy do you think Little Sheep is after knowing the combinations of those grasses? Please tell me by moving the arrow on the board." After the participant had moved the arrow, the experimenter said, "Your answer is X, isn't it? How do you get this answer?" Both the participants' answers and their explanations were recorded. In the experiment, participants were allowed to place the arrow between whole numbers. If the arrow was placed around the middle point of two whole numbers, the experimenter asked the participant if the answer was is the middle of the two numbers. If the participant answered, "Yes,", the experimenter recorded the answer as x.5. If the participant answered, "No," or the arrow was not around the middle point, the experimenter intended to take out a ruler to measure and record the answer with one decimal place, however, this last situation did not occur in the trials.

At the end of the pretest phase, Little Sheep thanked the participant for making predictions for her and said goodbye to the participant.

3.2.4.2 The training phase

At the beginning of the training phase, the experimenter gave the following instruction:

> Let us begin the second activity. In this activity, I will tell you whether your estimation is correct or not. Please pay attention to the correct answers, and try to think about why your answer is correct or incorrect. If your answer is correct, I will give you 2 points. If your answer is greater than or smaller than the correct answer by 1, I will give you 1 point, because the task is really difficult, and you are very smart to get your answer so close. If you get high marks in this test, I will give you a present. So please do your best when you make your estimation and you may try to think about previous questions and answers when you try to make an estimation.

The experimenter asked if there were any questions. If not, the experimenter introduced a new hand puppet to the participant:

> This is Little Horse. Just like Little Sheep, he likes delicious green grass and hates bitter brown grass. Now it is Little Horse's turn to draw some grass from the bag. You need to do estimations similar to the previous activity. After you have made an estimation, Little Horse will tell you whether your answer is correct or not.

As in the pretest, after confirming the correct number of delicious and bitter grasses identified, the participants were asked to indicate their answer using the Answer Board. They were also asked to explain how they arrived at a certain answer. After that, Little Horse told the participant the correct answer, the correct answer was also shown on the PowerPoint slide, and the experimenter moved the arrow of the Answer Board to the correct answer. The experimenter then recorded the participants answers.

At the end of the training phase, Little Horse thanked the participant for trying so hard and said goodbye to the participant. The experimenter told the participant the number of points achieved and gave the present to the participant after the third activity was over.

3.2.4.3 The posttest

At the beginning of the posttest, the experimenter gave the following instruction:

> Very good, I can see that you have tried very hard to do the estimations and to explain your answers. Now we will begin our third activity. Little Pig knows that you are able to perform estimations and because she thinks that you are

very smart, she wants to learn the method of doing estimations from you. Please be her teacher and tell her what you think when she asks you questions.

In the posttest, children first performed estimations following the same procedure as in the pretest. After the children had offered their explanations for questions 2 and 4, Little Pig came forward to give counter suggestions.

Question 1: 5 green grasses out of 10 grasses, probability value 0.5

Question 2: 3 green grasses out of 6 grasses, probability value 0.5

If a participant's answer to question 2 was within +/- 1 of the answer to question 1, Little Pig said:

> In the previous bag, there were five delicious grasses and five bitter grasses, and your answer was X. In this bag, there were three delicious grasses and three bitter grasses. There were more delicious grasses last time. Why is your answer so similar to your previous answer?

If a participant's answer was not within +/- 1 of the answer to question 1, Little Pig said:

In the previous bag, there were five delicious grasses and five bitter grasses, and your answer was X. In this bag, there were three delicious grasses and three bitter grasses. The number of delicious grasses and bitter grasses are the same both times. Why is your answer this time so different from your previous answer?

Question 3: 5 green grasses out of 9 grasses, probability value 0.56

Question 4: 3 green grasses out of 4 grasses, probability value 0.75

If a participant's answer to question 4 was greater than that of question 3, Little Pig said:

In the previous bag, there were five delicious grasses and four bitter grasses, and your answer was X. In this bag, there were three delicious grasses and one bitter grass. There are less delicious grasses this time, why is your answer greater than last time?

If a participant's answer to question 4 was smaller than or equal to the answer to question 3, Little Pig said:

In the previous bag, there were five delicious grasses and four bitter grasses, and your answer was X. In this bag, there were three delicious grasses and one bitter grass. There are less bitter grasses this time, why is your answer smaller than (or equal to) last time?

After completing question 5, Little Pig asked the following questions:

1. Do you think about the number of delicious grasses when you make your estimation?

2. Do you think about the number of bitter grasses when you make your estimation?

3. Do you think about the total number of grasses when you make your estimation?

4. Do you think about anything else?

5. You say you think about XX, YY, and ZZ, so after looking at XX, YY, and ZZ, how do you know the answer?

3.2.5 Analysis

3.2.5.1 Behavioural performance

The accuracy score is the absolute value of the difference between the correct answer and the participant's estimation. Correct estimation results in zero. The absolute distance between the correct answers and the estimations should be a more sensitive measure of children's ability than the dichotomous classification of "correct" and "incorrect," and subtle improvement in estimation performance should be detected more easily. The lower the accuracy score, the higher the accuracy in the probability estimation. ANOVA of the overall accuracy score for each participant was performed. The mean accuracy scores of the pretest and the posttest were calculated and were compared to see if there was any improvement.

Bryant (1974) suggested that, when the part-whole concept is not yet developed, using relations of "greater than," "smaller than," and "equal to" to compare constituent parts could be the first step in the process of quantification of fractions. In the probability estimation task of this study, "half" was also a very useful concept. If the number of both types of grasses were equal, the answer must be 5. If the number of green grasses was more than brown grasses (or the number of green grasses was more than half of the total number of grasses), then the answer must be greater than 5. If there were more brown grasses than green grasses, then the answer must be smaller than 5. By using this "half rule," accuracy of estimation could be greatly improved.

The behavioural half rule score (BHalf score) of each participant was obtained by counting the number of trials that adhered to the half rule in all three phases. For example, if there were 2 green grasses and 1 brown grass, the correct answer was 6.7.

If the participant's response was 9, though the response was not correct, the participant had adhered to the half rule: there was more green grass than brown grass and the participant's answer was greater than 5. If the participant's response was 5, though it was more accurate than the response of 9, this answer violated the half rule: 5 could only be used when both types of grasses were equal in number.

3.2.5.2 Performance that demonstrates explicit understanding

In Siegler and Stern's (1998) study, they classified children's computational strategies into four types, and studied how children progress from lower strategies to higher strategies. In this experiment, the verbal explanations provided by the participants in each trial were categorized into different categories.

The classification system used in the pilot test was amended after examination of data obtained in the main study. Originally, the following seven types of explanation were created, based on the literature: Implicit computation, Numerator only, Numerator and denominator only, Subtraction, Half rule, Others (Idiosyncratic), and Others (Correct). Two categories were eliminated from the old system: Numerator Only and Numerator and denominator only. They were designed into the original system because research suggested that children would only consider the numerator when they made probability decisions. Therefore, the researcher believed that it was meaningful to include these two categories for investigating the issue. However, when the researcher tried to classify the experimental data, it was found that these two categories often overlapped with other types of explanation, and therefore the researcher decided to restructure the classification system. The new classification system aimed at classifying the different types of computations that were reflected by the explanations. The categories "Numerator only" and "Numerator and denominator only" were removed, because they were not computation methods. The original categories "Subtraction" and "Others (Correct)" were renamed "Comparison" and "Division/Fraction" respectively, in order to represent the essence of these computational methods more accurately. A new category, "Plus minus," was created after the examination of empirical data.

To assess the interrater reliability, transcripts and video recordings of 10 participants (41.7% of all of the trials) were given to a second rater to independently perform classification. The first rater was the researcher. The second rater was a PhD student. The interrater reliability was evaluated using Cohen's Kappa. With $\kappa = .794$, the degree of agreement was substantial (strength of agreement based on Landis and Koch, 1977).

In the new classification system, explanations provided by participants were classified into seven categories:

1) **Implicit computation (I)**: The participants said that they did not know, or the explanation fulfilled **all** of the following requirements:

    i. Participant did not mention the relationship between the quantity of grass and the answer.

    ii. Participant did not comment on the quantity of grass.

    iii. Participant did not make any comparison.

2) **Last question (L)**: The participant's explanation was based on information from previous question(s).

3) **Comparison (C)**: The participant explained by comparing the number of green and brown grasses to identify which type occurred more or less times.

4) **Plus minus (P)**: The participant assigned a score to each green or each brown grass image and the answer was obtained by adding or subtracting these scores. The score carried by brown grass must be used for deduction in the computation process.

5) **Half rule (H)**:

    i. ***If the number of green grasses was equal to the number of brown***

grasses: The participant's response must be 5, and the explanation indicated that both types of grasses were the same.

ii. ***If the number of green grasses was more than the number of brown grasses:*** The participant's response must be greater than 5, and the explanation indicated that the number of green grasses was more than the number of brown grasses: therefore, the answer should be greater than 5.

iii. ***If the number of green grasses was less than the number of brown grasses:*** The participant's response must be less than 5, and the explanation indicated that there were more green grasses than brown grasses; therefore, the answer should be greater than 5.

6) **Division/Fraction (F)**: The participant explained correctly using the concept of division or fractions.

7) **Others—Idiosyncratic (O)**: The participant provided a computation method or relationship between quantity of grass and the response that could not be classified within the above categories.

Like the explanation analysis of the block-balancing task, the distribution was analyzed from two directions: the frequency of occurrence among the trials in each age group, and the number of participants who had used a certain type of explanation in each age group. The accuracy score of each explanation type was calculated in order to see which type of explanation gave a more accurate answer. The conceptual understanding of participants reflected by these explanation types was also analyzed.

In the posttest phase of Study Two, the experimenter provided counter suggestions to the participants based on the answers they provided in the previous two questions. If the participants were able to produce correct estimations and they had full insight into their answers' correctness, they should not be misled by the counter suggestions. The participants' responses to counter suggestions was analyzed in order to see how well they understood their own estimations.

At the end of the experiment, the experimenter asked some general questions about the computation method used in the probability task. If the developmental direction of this task followed the top-down direction, participants should perform better in general questions about computation, because this was the starting point of the development. Explicit knowledge would be the guide for carrying out computation and for getting the answer. The development should happen in this way:

1) The participants' acquisition of explicit knowledge about the general principle of computation, which was represented at E2/3 level, becomes available to conscious access and can be verbalized. 2) The participants with a lower level representation of knowledge (E1 or level-I) cannot verbalize, intermediate steps are unavailable to conscious access, but computation is carried out automatically and efficiently. This is what Anderson found (1982) when he studied how students learn to solve geometric problems using postulates. He found that in the first stage, that is the declarative stage, the participants first memorized declarative knowledge. Upon execution, the participants might verbally rehearse the explicit knowledge that they needed to interpret the explicit knowledge and turn the knowledge into operation. In the second stage, declarative knowledge was compiled into procedures. After composition and proceduralization, procedures became encapsulated, and intermediate steps were unavailable to conscious access. The responses provided by participants in response to general questions and their explanations provided in these experimental trials were compared, in order to see which one was more advanced.

3.2.5.3 Comparison of behavioural performance and performance that demonstrates explicit understanding

The first comparison concerns the trials in which Explanation I was provided. In trials where no explicit computation method was provided by the participant, the answer provided by the participant should be generated by intuitive estimation. The computational process of intuitive estimation is not available for conscious access; therefore, the representation of intuitive estimation is not represented explicitly. If there is only one kind of representation for knowledge, namely explicit knowledge, or explicit knowledge replaces implicit knowledge in the process of development, then, among trials of Explanation I, the accuracy score should be the same in all of the age groups, because no explicit knowledge was used in this kind of trial. Comparisons were made to test whether the accuracy was the same in all of the age groups for this kind of trial.

The second aim was to compare the pace of development between behavioural performance and explicit verbal explanation. If these two kinds of performance were sustained by the same kind of representation, then when one kind of performance improves, the other kind of performance should also improve. If there were differences between the two, then it is more reasonable to take the position that the two are sustained by different kinds of representation. If there were different kinds of representation and the developmental direction was top-down, then improvement in verbal explanation should precede behavioural improvement, because explicit knowledge is the precondition for improvement in implicit knowledge. To see whether the difference in performance exists, pretest and posttest performance of both kinds of performance were compared.

As mentioned before, the half rule may be a stepping-stone for children to develop correct concepts about probability. It would be meaningful to find out whether there was a difference in performance of the behavioural and verbal versions of the half rule.

3.3 Notes about feedback in the two experiments

In both Study One and Study Two, the participants received feedback. In the block-balancing experiment, the participants received feedback about whether the block fell or was successfully balanced. In the probability-estimation task, feedback was provided in the training phase in the form of a correct answer. This correct answer could not be viewed as equivalent to providing explicit knowledge. In the experiment, it was repeatedly found that participants were puzzled by the correct answers. They often asked the experimenter why the correct answer was correct, because the correct answer itself could not tell the participant how to perform the computation. The implication of the answer was completely subject to the children's interpretation, and this was affected by the participants' conception of the task. The same was also true for the feedback in the block-balancing task, because Karmiloff-Smith (1992) pointed out that:

> In the present microdomain, when a weighted block balances off-, this represents *positive* feedback for the younger children because it meets their goal. However, the very same stimulus represents *negative* feedback (a balanced block) for older children holding the geometric- theory. Similarly, when a block placed off- falls, this represents negative feedback for the younger children, whereas the same stimulus represents positive feedback for the somewhat older children because a failed off- attempt confirms their geometric- theory (p.87).

3.4 The pilot study and amendments made after the pilot study

Pilot tests of Study One and Study Two were conducted in October, 2007. A total of 10 children participated in the pilot study (three of them undertook the pilot test of Study One, whereas seven of them undertook the pilot test of Study Two). Five of these children came from a tutorial school in Wanchai, two were children of the researcher's schoolmates, and another two went to the Sunday school of the researcher's church. One of the participants was the child of a university professor. The purpose of the pilot study was to find out whether the experimental design achieved the purpose of the study, and whether the intended findings can be elicited. Another aim was to test whether the experimental design was feasible and whether the tasks were suitable for the selected age groups. Experience gained in the pilot study was used to improve the experimental design of this study.

The three participants who participated in the pilot study of Study One were one from each age group. None of the participants showed difficulty in understanding the instructions. The youngest participant demonstrated the implicit geometric-centre theory by his persistent central-area placement, but it is unclear whether the geometric-centre theory was represented at E1, since he did not mention anything about the middle point. With a view to better differentiating level-E1 representation from level-I representation among the children, it was decided to add the Prediction Task into the main study. The free play period was reduced to five minutes, because the experimenter observed that the children began to use the blocks to do things other than balancing if the free play time went on for too long. It took more than half an hour for the whole experiment to be completed in the pilot test. The experiment for

Study One was adjusted to last about 30-35 minutes, after amendments to the pilot experiment. Three questions to gauge the children's awareness of the geometric-centre theory were added at the end of the experiment, so as to understand more about the metacognitive status of the geometric-centre theory.

Seven children participated in the pilot test of Study Two. There were three participants in the youngest age group, one participant in the middle age group, and three participants in the eldest age group. The result of the pilot study showed that even the youngest participant was able to tackle the task and the score obtained was similar to the score of some of the older participants. After the pilot test of the first participant, the stimuli set was reconstructed to shorten the time of the experiment. In the pilot test of the second participant, it was found that the original method of stimuli presentation was too time consuming and the participants became distracted. Starting from the pilot test of the third participant, a notebook computer was used for presenting the stimuli. The remaining pilot tests lasted around 30–35 minutes.

The story told by the experimenter in this pilot test was slightly modified. In the pilot test, children were told that Little Sheep or Little Horse went to visit different grasslands in the Sun Kingdom. The animals were unable to distinguish delicious grass from bitter grass and they could not remember which grass they had just eaten. The grass in the Sun Kingdom were very special, after it was eaten it immediately grew again and looked the same as before. These instructions aimed to create a scenario of random draw with replacement. However, this was too complicated for the children to grasp. The experimenter had to remind participants about these instructions throughout the experiment. Therefore, in the main study, the scenario of random draw was created with another set of instructions. The participants were told that Little Sheep or Little Horse had to take grass from a transparent plastic bag with their eyes closed. Such a scenario was easier for the children to understand and remember, and participants in the main study showed no problem in understanding the constructed scenario.

CHAPTER FOUR: RESULTS AND DISCUSSION OF STUDY ONE

4.1 Outline

The performance of the block-balancing task will be reported from two directions: behavioural performance and performance that demonstrates explicit understanding.

In this study, the behavioural performance in the block-balancing task was measured in three ways: success in block balancing, initial placement of the blocks, and the placement area of the blocks. First, the basic dimensions for measuring behavioural performance will be introduced, then the two behavioural patterns, namely the behavioural pattern of geometric-centre theory (BGeo) and the behavioural pattern that reflects the naïve version of the law of torque (BTorque), will be discussed. The development of these two patterns will be compared with the development of the explicit concepts that are involved in the block-balancing task.

For the performance that demonstrates explicit understanding, the Prediction Task will be discussed first, then the explanations provided by participants will be classified into different types, and finally, the characteristics and developmental trends of the explanations will be discussed.

The behavioural performance and the explicit understanding of the participants will then be compared to see whether the distinction between implicit and explicit representation can be observed. The questions that will be investigated are whether implicit geometric-centre theory exists in E1 form, and whether development occurs from the top-down direction or the bottom-up direction.

4.1.1 Notes on the participant code

In this experiment, the performance was studied individually. To protect the privacy of the participants, they are assigned a participant code in the report of the results. An example of a participant code is "AF1." The first letter indicates age group: "A" for age 4 to 5 years old, "B" for age 6 to 7 years old, and "C" for age 8 to 9 years old. The second letter indicates gender: "F" for female and "M" for male. The last digit ranges from 1 to 4. The participant code "AF1" thus means "the first female participant of the 4- to 5-year-old age group."

4.2 Behavioural performance

4.2.1 Success Score

Success Scores of symmetrical blocks (Blocks A, B, C, and D) and asymmetrical blocks (Blocks E, F, G, and H) were obtained by counting the number of successfully balanced blocks in each group.

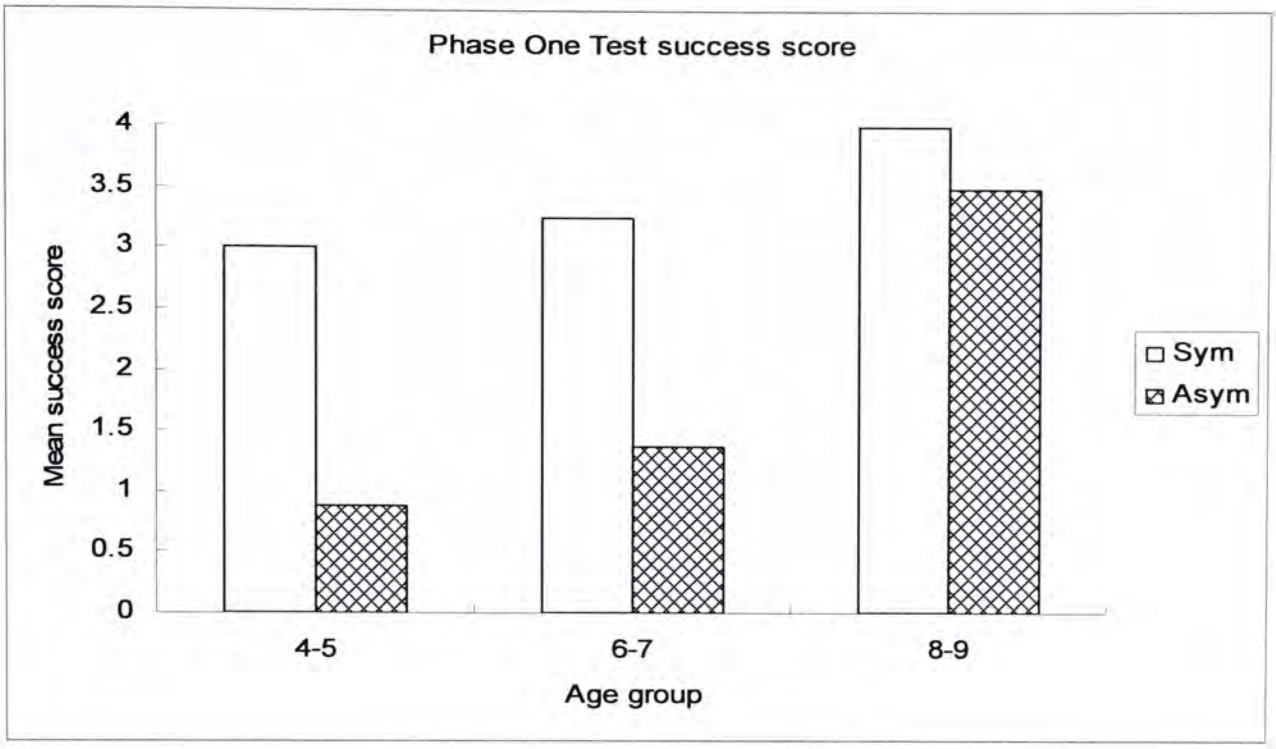

Fig. 4.1 Success Score of Phase One and Phase Two

The Success Score was analyzed using ANOVA. The within-subjects factors were block type (Symmetrical and Asymmetrical) and phase (Phase One and Phase Two), and the between-subjects factors were age group (Age groups 4-5, 6-7, and 8-9) and gender (Male and Female).

The main effect of age group is significant, $F(2, 18) = 9.405$, $p < .05$, Partial Eta Squared = .51, representing a large effect. The main effect of gender is not significant, $F(1, 18) = 2.01$, $p > .05$. The effect of phase factor is not significant, $F(1, 18) = 2.65$, $p > .05$. The main effect of block type is significant, $F(1, 18) = 40.933$, $p < .01$, Partial Eta Squared = .695, which is a large effect.

None of the interactions is significant except the interaction between block type and age, $F(2, 18) = 7.067$, $p < .01$, Partial Eta Squared = .44. The Bonferroni adjusted pairwise comparisons show that for asymmetrical blocks, the mean differences between age groups 4 to 5 years old and 8 to 9 years old, and 6 to 7 years old and 8 to 9 years old are significant. All other mean differences are not significant.

In Phase Two, all participants in the 8- to 9-year-old age group could balance all of the symmetrical and asymmetrical blocks used in the experiment.

4.2.2 Initial Middle Placement Score

Initial Middle Placement Scores of symmetrical and asymmetrical blocks were obtained by counting the number of blocks being initially placed at the geometric-centre.

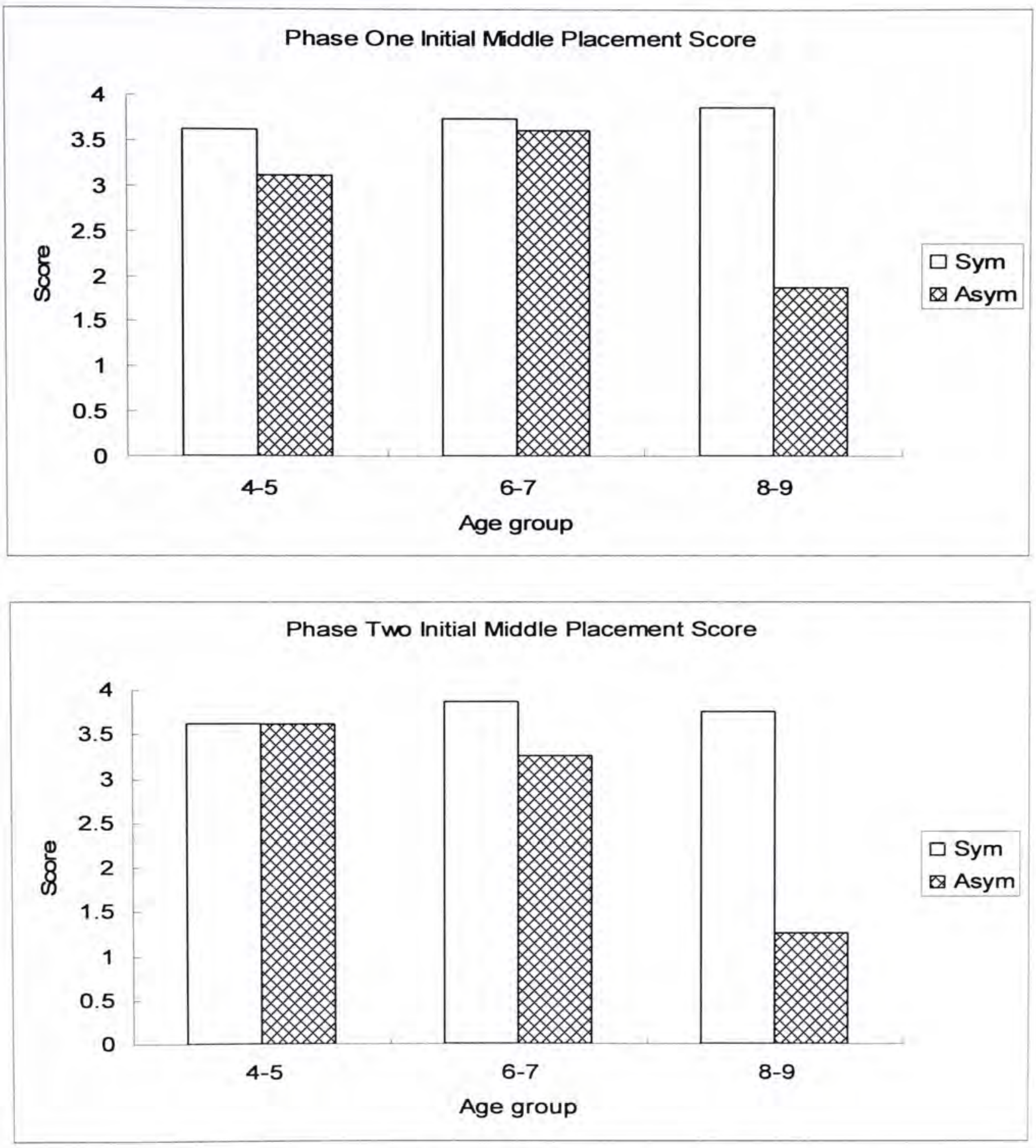

Fig. 4.2 Initial Middle Placement Score of Phase One and Phase Two

The Initial Middle Placement Score was analyzed using ANOVA. The within-subjects factors were block type (Symmetrical and Asymmetrical) and phase (Phase One and Phase Two), and the between–subjects factors were age group (Age groups 4-5, 6-7, and 8-9) and gender (Male and Female).

The main effect of age group is significant, $F(2, 18) = 14.103$, $p < .01$, Partial Eta Squared = .61. This finding is different from the study performed by Krist et al. (2005). In their study, the main effect of age was insignificant, and only the interaction between age and block type was significant.

The main effect of gender is insignificant, $F(1, 18) = .557$, $p > .05$, and the effect of phase is not significant, $F(1, 18) = .421$, $p > .05$. However, the main effect of block type is significant, $F(1, 18) = 30.584$, $p < .01$, Partial Eta Squared = .630.

The only significant interaction effect is between block type and age, $F(2, 18) = 14.496$, $p < .01$, Partial Eta Squared = .617. The Bonferroni adjusted pairwise comparisons show that for asymmetrical blocks, the mean differences between age groups 4 to 5 years old and 8 to 9 years old, and 6 to 7 years old and 8 to 9 years old are significant. All other mean differences are not significant.

4.2.3 Geometric Centre Area Placement Score (GCP Score)

Geometric Centre Area Placement Scores (GCP Scores) of symmetrical and asymmetrical blocks were obtained by counting the number of blocks placed in the geometric-centre area only, throughout the trial, for that block.

The GCP Score was also analyzed using ANOVA. The within-subjects factors were block type (Symmetrical and Asymmetrical) and phase (Phase One and Phase Two), and the between-subjects factors were age group (Age groups 4-5, 6-7, and 8-9) and gender (Male and Female).

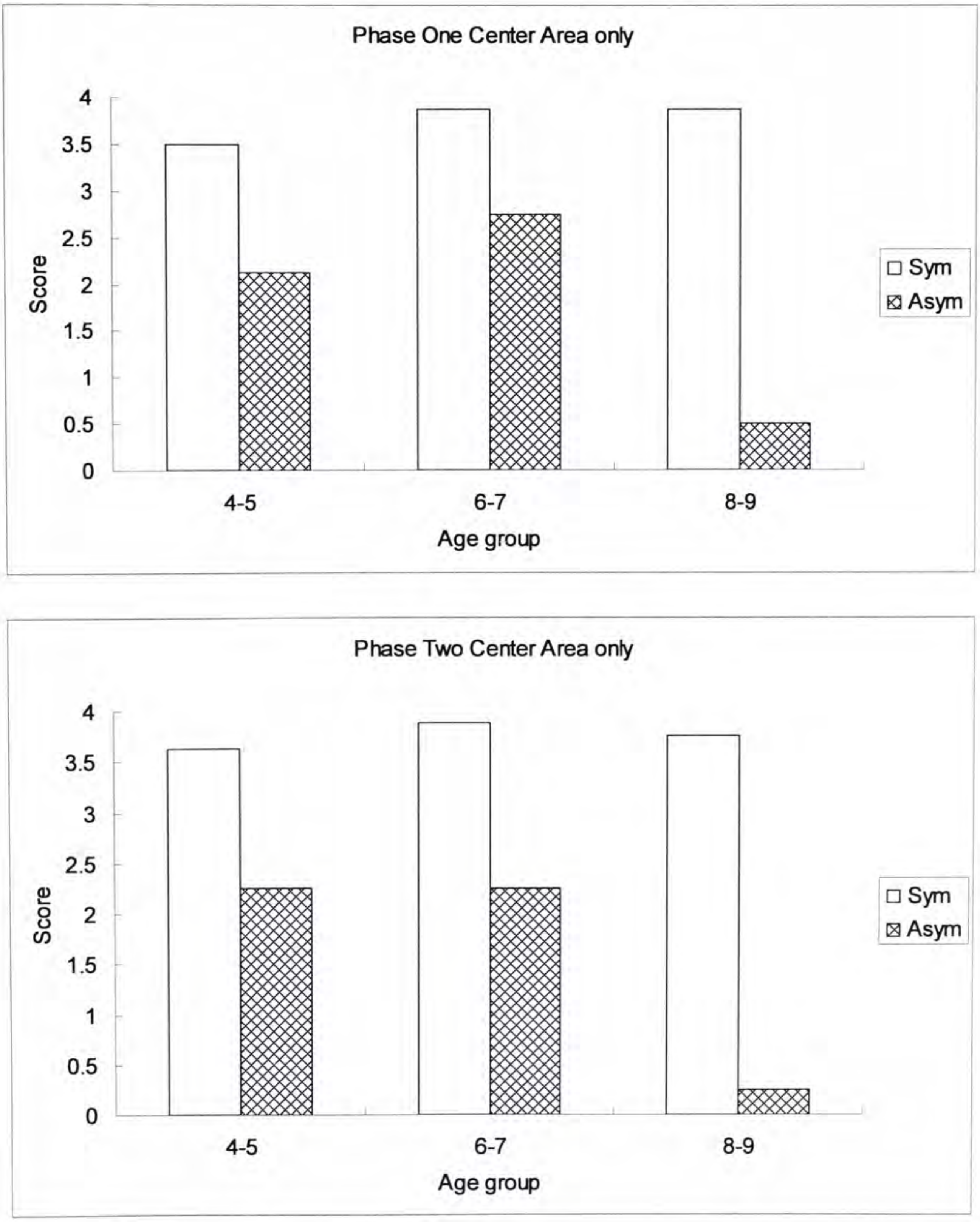

Fig. 4.3 GCP Score of Phase One and Phase Two

The main effect of age group is significant, $F(2, 18) = 3.895$, $p < .05$, Partial Eta Squared = .302. The effect of gender is not significant, $F(1, 18) = .324$, $p > .05$. The within-subjects factor phase is not significant, $F(1, 18) = .728$, $p > .05$, while block type is significant, $F(1, 18) = 64.059$, $p < .01$, Partial Eta Squared = .781.

Again, the only significant interaction is between block type and age group, $F(2, 18) = 7.118$, $p < .05$, Partial Eta Squared = .442. The Bonferroni adjusted pairwise comparisons show that the only significant mean difference is between the means of the asymmetrical block scores of the 6- to 7-year-old and 8- to 9-year-old age groups.

4.2.4 Comparison between the three behavioural measures

A good performance in the Success Score measurement requires high marks for both symmetrical and asymmetrical blocks. Good performance in the Initial Middle Placement Score and GCP Score measurements requires high marks for symmetrical blocks and low marks for asymmetrical blocks. By comparing Figures 4.1, 4.2, and 4.3, two observations can be made:

1. ***The development trends reflected by the measures are different:*** Looking at the Success Score, a linear increase of score with age is found for both phases. This agrees with Krist et al.'s (2005) finding. However, this linear improvement trend

cannot be observed in the Middle Placement Score of Phase One and the GCP Score for in the two phases.

2. ***The GCP Score is more sensitive than the Middle Placement Score:*** Participants of the 8- to 9-year-old age group could balance all of the asymmetrical blocks in Phase Two. Comparing the Middle Placement Score and the GCP Score, the GCP Score is more sensitive in detecting this change because the score is markedly lower for the GCP Score than for the Middle Placement Score.

The ANOVAs of the three measures all agree that the only significant interaction is the interaction between block type and age group, and they all agree that the main effect of age group and block type is significant. Considering the interaction effect between block type and age group, all three measures agree that for asymmetrical blocks, the difference between the 6 to 7 year old and 8 to 9 year old groups is significant.

The difference in performance with the asymmetrical blocks between the 4 to 5 year old and 8 to 9 year old groups, is significant for Success Score and Initial Middle Placement Score, but it is not significant for the GCP Score.

To prove that there is a U-shaped performance curve for asymmetrical blocks across age groups, the performance of the 6 to 7 year olds should be significantly worse than the performance of both the 4 to 5 year olds and the 8 to 9 year olds. However, the difference of asymmetrical block performance between 4 to 5 year olds and the 6 to 7 year olds was not significantly different in the three age groups. Thus, there is no support for the existence of a U-shaped performance curve in this experiment.

4.2.5 Behavioural pattern

4.2.5.1 Behavioural pattern of the geometric-centre theory (BGeo)

Test Phases One and Two were classified as exhibiting a behavioural pattern of geometric-centre theory (BGeo), if the GCP Score for asymmetrical blocks in that test phase was greater than or was equal to 3.

	Age 4-5 (A)								Age 6-7 (B)								Age 8-9 (C)							
	AF1	AF2	AF3	AF4	AM1	AM2	AM3	AM4	BF1	BF2	BF3	BF4	BM1	BM2	BM3	BM4	CF1	CF2	CF3	CF4	CM1	CM2	CM3	CM4
Phase 1	G	-	G	-	-	-	G	G	-	G	G	G	G	G	-	-	-	G	-	-	-	-	-	-
Phase 2	G	-	G	G	G	-	G	G	-	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-

Table 4.1 Distribution of BGeo by individual participants, "G" indicates the behavioural pattern of geometric-centre theory

	Age 4-5	*Age 6-7*	*Age 8-9*
Phase One	4	5	1
Phase Two	6	5	0

Table 4.2 Distribution of BGeo by age group (8 participants per age group)

It can be seen that the BGeo was the prevailing behavioural pattern among the groups of 4 to 5 year olds and 6 to 7 year olds. In Phase Two, 6 of the 8 participants in the group of 4 to 5 year olds, and 5 of the 8 participants in the group of 6 to 7 year olds exhibited the BGeo. None of the participants from the group of 8 to 9 year olds exhibited this behavioural pattern in Phase Two.

In the 4- to 5-year-old group, 2 participants did not exhibit the BGeo in Phase One, but developed this pattern in Phase Two. In the 8- to 9-year-old group, one participant exhibited the BGeo, but abandoned this pattern in Phase Two. The transcript of this participant reveals that the participant discovered blocks can be balanced off centre when she successfully balanced the last block of Phase One (Block H, asymmetrical block). She was surprised by the result:

> *CF2: This thing is really strange. Because... cannot... put that... that is cannot balance it at centre, must put it near the side a little bit, so I don't know why...*
>
> *(Participant CF2, Phase One, Trial of block H)*

4.2.5.2 Behavioural pattern reflecting the naïve version of the law of torque (BTorque)

Phase One and Phase Two tests were classified as exhibiting BTorque if 3 or more of the trials for asymmetrical blocks were successfully balanced and the initial placement position of the block was on the correct side.

	Age 4-5 (A)								Age 6-7 (B)								Age 8-9 (C)							
	AF1	AF2	AF3	AF4	AM1	AM2	AM3	AM4	BF1	BF2	BF3	BF4	BM1	BM2	BM3	BM4	CF1	CF2	CF3	CF4	CM1	CM2	CM3	CM4
Phase 1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	T
Phase 2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	T	T	-	T	-	T

Table 4.3 Distribution of BTorque by individual participants, "T" indicates the behavioural pattern of the naïve version of the law of torque.

	Age 4-5	*Age 6-7*	*Age 8-9*
Phase One	0	0	2
Phase Two	0	1	4

Table 4.4 Distribution of BTorque by age group (8 participants per age group)

None of the participants in the group of 4 to 6 year olds exhibited BTorque. Except for one participant, who came from the group of 6 to 7 year olds and exhibited BTorque in Phase Two, all of the other participants who exhibited BTorque came from the group of 8 to 9 year olds. Half of the participants in the group of 8 to 9 year olds exhibited BTorque in Phase Two. The result suggests that BTorque appears rather late developmentally.

4.3 Performance that demonstrates explicit understanding

Performance of explicit understanding will be approached from two directions: one is participants' performance in the Prediction Task, the other is the explanations provided by participants during the Phase One and Phase Two tests.

4.3.1 Prediction Task

In the Prediction Task, participants were required to classify all of the blocks into two groups: "can be balanced" and "cannot be balanced." Because all of the stimuli blocks could be balanced, a block that was classified as "can be balanced" scored one point in this task. The score of the task was analyzed using ANOVA. The between-subjects factors were gender (Male and Female), and age group (Age groups 4-5, 6-7, and 8-9). The within-subjects factor was block type (Symmetrical and Asymmetrical).

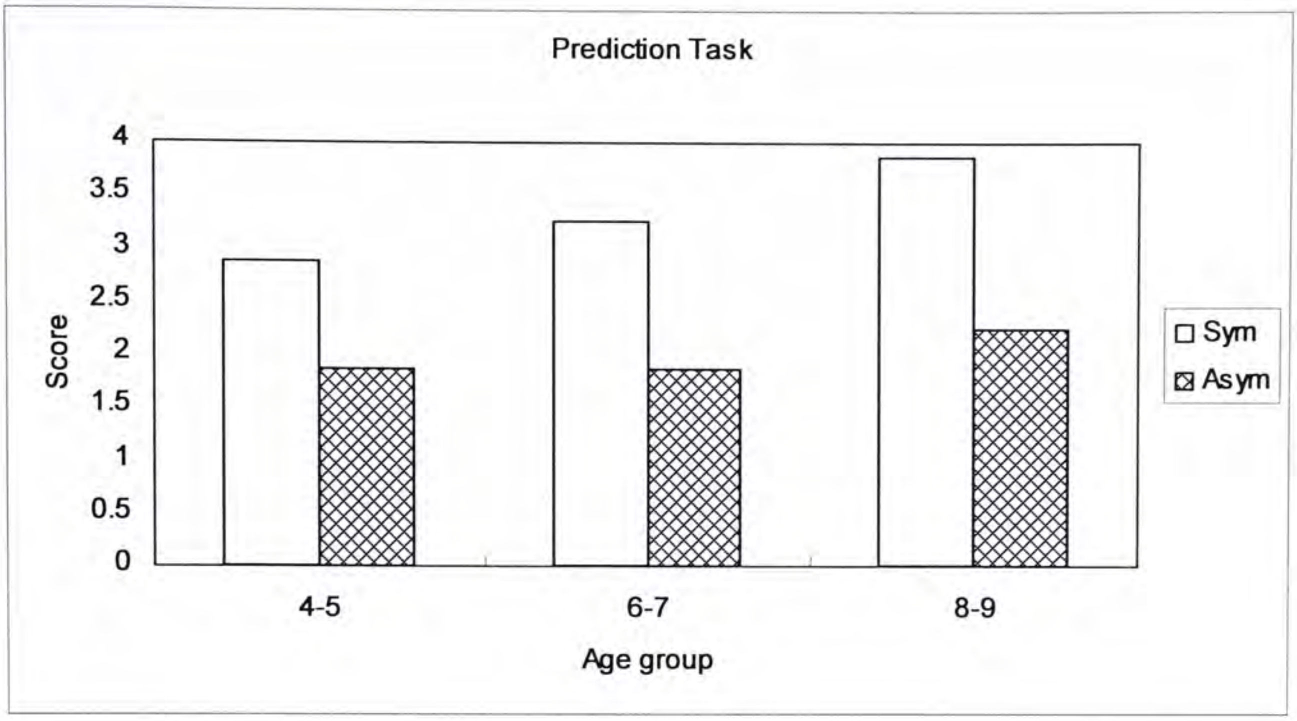

Fig. 4.4 Score of the Prediction Task

The only significant main effect is block type, $F(1, 18) = 24.381$, $p < .01$, Partial Eta Squared = .575. The main effect of age group is not significant, $F(2, 18) = .977$, $p > .05$. This is different from the result of behavioural measures, where for all three behavioural measures, the effect of age group is significant. The effect of gender is not significant, $F(1, 18) = .362$, $p > .05$. The effect of interaction between age group and block type is significant in all behavioural measures, but is not significant in the Prediction Task, $F(2, 18) = .452$, $p > .05$. The only significant interaction for the Prediction Task is the three-way interaction between block type, age group, and gender, $F(2, 18) = 5.167$, $p < .05$, Partial Eta Squared = .365. This finding also contradicts the analysis of behavioural measures, because the effect of this interaction is not significant in all three behavioural measures. Bonferroni adjusted pairwise comparisons show that for the group of 8 to 9 year olds, for asymmetrical blocks, the mean between male and female is different.

	CF1	CF2	CF3	CF4	CM1	CM2	CM3	CM4
No. Asym blocks success: Phase One	4	1	4	4	4	3	4	4
Prediction Score: Symmetrical	3	4	4	4	4	4	4	4
Prediction Score: Asymmetrical	3	4	4	4	0	3	0	0
No. Asym blocks success: Phase Two	4	4	4	4	4	4	4	4

Table 4.5 Prediction Score for the 8- to 9-year-old group (Asym = Asymmetrical)

Referring to Table 4.5, it is surprising to find that participants CM1, CM3, and CM4 predicted that all of the asymmetrical blocks could not be balanced. However, all three participants had successfully balanced all of the asymmetrical blocks in test Phase One, which preceded the Prediction Task, and in test Phase Two, which immediately followed the Prediction Task.

Symmetrical Block				Asymmetrical Block			
Subject	Block	Phase 2 success	Exp T in Phase 1	Subject	Block	Phase 2 success	Exp T in Phase 1
AF4	Block B	N	N	BM3	Block E, F	Y	Y
AM1	Block D	Y	N	CF1	Block F	Y	Y
AM4	Block D	Y	N	CM1	Block E, F, G, H	Y	Y
BF4	Block D	N	N	CM2	Block H	Y	Y
BM1	Block D	Y	N	CM3	Block E, F, G, H	Y	Y
BM2	Block D	Y	N	CM4	Block E, F, G, H	Y	Y
CF1	Block C	Y	Y				

Table 4.6 Blocks being successfully balanced in Phase One but predicted as could not be balanced. (Exp T = Explanation type T)

Other than the above mentioned three surprising cases, there were other cases of contradiction between the Phase One behavioural performance and the Prediction

Task, which were reported in Table 4.6. The patterns of contradiction for symmetrical blocks and asymmetrical blocks were quite different. On the "Symmetric Block" side of Table 4.6, participants were from the younger group as compared with those on the "Asymmetrical Block" side. On the "Symmetric Block" side, each participant only made one contradiction for symmetrical blocks, and no participant would produce contradiction for all of the symmetrical blocks. The majority of these contradictions occurred for block D. Though these participants had successfully balanced the symmetrical blocks in test Phase One, they did not always repeat this success in test Phase Two. Only one of these participants provided explanation type T in test Phase One (details of explanation type T will be discussed in the next section). However, for asymmetrical blocks, no single type of block was responsible for the majority of the contradictions. Three participants produced contradictions for all of the asymmetrical blocks. All of the participants were capable of balancing the asymmetrical blocks again in Phase Two, and all of them had provided explanation T in test Phase One.

In the group of symmetrical blocks, block D was the most common block that was associated with contradiction. Block D and block H looked visually the same, but block H was an asymmetrical block. Block H was the last stimulus in Phase One, and all of the participants who predicted block D could not be balanced in Table 4.10 failed to balance block H in Phase One. They all predicted that neither block D nor block H could be balanced in the Prediction Task. It seems that these participants' prediction about block D was due to the recent failure of block H.

It is easier to dismiss the contradictions that occurred with the symmetrical blocks as memory failure. However, the anomaly with the asymmetrical blocks cannot be easily dismissed in the same way. An appropriate explanation is needed for the contradictions that were expressed by the participants regarding the asymmetrical blocks. This interesting finding will be discussed later in this chapter.

4.3.1.1 Comparing the Prediction Score with the Success Score

Comparing Figure 4.8 (Prediction Score) with Figure 4.1 (Success Score), the developmental trends reflected by the two figures are markedly different. The Success Score increases with age and ANOVA shows the effect of age group is significant. For the Prediction Score, symmetrical blocks show a similar increasing trend, but for asymmetrical blocks, the scores for the three age groups are very similar, and ANOVA shows that the effect of age group is not significant.

The Success Score reflects behavioural success and the Prediction Score reflects performance of conscious inference. If behavioural performance and explicit understanding are sustained by the same representation, it seems difficult to understand why the discrepancy between Success Score and Prediction Score occurs. It seems more reasonable to interpret the discrepancy by proposing that behavioural success is sustained by knowledge that can only be used in performing the block-balancing behaviour, because the knowledge cannot be used for making predictions. If this is the case, then the knowledge that contributes to behavioural success should be represented at level-I; and the performance of the Prediction Task should be caused by knowledge represented at a level other than level-I.

4.3.2 Explanation types

From the perspective of the RR model, the explanations provided by the participants should reflect knowledge represented at E2/3 level, because this is the only level that can be reported verbally. The verbal explanations provided by participants in each trial were classified into eight categories, namely: (1) N: No explanation; (2) P1: Physical properties 1; (3) P2: Physical properties 2; (4) P3: Physical properties 3; (5) G: Geometric-centre theory; (6) B: Balance point related; (7) T: The naïve version of the law of torque; and (8) O: Others—Idiosyncratic. For the classification requirements, please refer to section 3.1.5.2.

The distribution of explanation types is shown in Figures 4.11, 4.12, and 4.13. As reported in section 3.1.5.2, the interrater agreement on the explanation types is almost perfect.

	Symmetrical Block				Asymmetrical Block		
	Age Group				Age Group		
	4-5	6-7	8-9		4-5	6-7	8-9
N	16 (25%)	12 (19%)	2 (3%)	N	9 (14%)	8 (13%)	3 (5%)
P1	8 (13%)	2 (3%)	0 (0%)	P1	9 (14%)	4 (6%)	0 (0%)
P2	15 (23%)	18 (28%)	22 (34%)	P2	32 (50%)	30 (47%)	6 (9%)
P3	14 (22%)	8 (13%)	0 (0%)	P3	6 (9%)	1 (2%)	0 (0%)
G	4 (6%)	9 (14%)	15 (23%)	G	0 (0%)	0 (0%)	0 (0%)
B	1 (2%)	0 (0%)	6 (9%)	B	6 (9%)	5 (8%)	5 (8%)
T	0 (0%)	9 (14%)	16 (25%)	T	0 (0%)	13 (20%)	47 (73%)
O	6 (9%)	6 (9%)	3 (5%)	O	2 (3%)	3 (5%)	3 (5%)

Table 4.7 Number and percentage (relative to each age group) of all explanations

	Age group							
	4-5		6-7		8-9		Total	
N	25	(20%)	20	(16%)	5	(4%)	50	(13%)
P1	17	(13%)	6	(5%)	0	(0%)	23	(6%)
P2	47	(37%)	48	(38%)	28	(22%)	123	(32%)
P3	20	(16%)	9	(7%)	0	(0%)	29	(8%)
G	4	(3%)	9	(7%)	15	(12%)	28	(7%)
B	7	(5%)	5	(4%)	11	(9%)	23	(6%)
T	0	(0%)	22	(17%)	63	(49%)	85	(22%)
O	8	(6%)	9	(7%)	6	(5%)	23	(6%)

Table 4.8 Total number and percentage of all explanations

	Age group		
	4-5	6-7	8-9
N	5	5	3
P1	6	4	0
P2	7	7	5
P3	6	3	0
G	3	3	5
B	3	2	3
T	0	2	8
O	3	3	2

Table 4.9 Number of participants in each age group who provided a specific type of explanation (8 participants in each age group)

The distribution of explanation types will be discussed from two directions: distribution among trials and distribution among individuals.

Considering the distribution among trials, the top three categories of explanations were P2 (32%), The naïve version of the law of torque (22%) and No explanation (13%). Added together, these top three categories of explanation were used in 67% of the trials.

Among 24 participants, the top three types of most frequently used explanation were as follows: P2 (19 participants, 79.2%), No explanation (13 participants, 54.2%), and Explanation involving the geometric centre (11 participants, 45.8%).

Explanation T (The naïve version of the law of torque) is the most correct type of explanation. "Most correct" means that the level of correctness is the highest when compared with other types of explanation in this study. It is unreasonable to expect children in this experiment to provide a completely correct explanation. None of the participants in the youngest age group could provide this type of explanation (Table 4.9), though there was one participant in this age group who was capable of balancing all of the blocks. The first appearance of explanation T was in the age group of 6 to 7 year olds, when 2 of the 8 participants gave this type of explanation. In the eldest age group, all of the participants provided this type of explanation at least once.

In terms of individual development across trials, explanation T was never the first type of explanation used by a participant. The first occurrences of T in all 10 participants were in the trials of asymmetrical blocks. Siegler and Jenkins (1989) found that more challenging problems stimulated discovery of a more advanced strategy. Asymmetrical blocks are more difficult to balance than symmetrical blocks; therefore, children were challenged to provide a more comprehensive explanation.

	Occurrence order (Counting from the trial T first occur)		
Subject	1	2	3
BM3	B	G	N
BM4	G	P2	
CF1	P2	N	G
CF2	B		
CF3			
CF4			
CM1	P2	G	
CM2	P2	O	
CM3	B	O	
CM4	P2	B	

Table 4.10 Types of explanation found after the first occurrence of Explanation T

Explanation T is correct for explaining all of the blocks in the experiment. However, after the first use of explanation T, not all participants shifted to using explanation T for the rest of the trials. Rather, of the 10 participants who had provided explanation T (Table 4.10), 8 still used other explanations in later trials. Moreover, 6 of them still used incorrect explanations (Explanation type P2, O, or N) in later trials. This finding aligns with Siegler and Jenkins' (1989) finding that a more advanced strategy did not completely replace less advanced strategies once it had been discovered; less advanced strategies were only phased out gradually.

Explanation type G is the explicit version of geometric-centre theory. All age groups had participants who provided this kind of explanation at least once (Table 4.9). All explanation G explained was used for explaining symmetrical blocks (Table 7). Explanation G never appeared in trials of the asymmetrical blocks. It should be noted that explanation G is correct for the symmetrical blocks, but incorrect for the asymmetrical blocks.

The explanations about physical properties (P1, P2, and P3) were all incorrect explanations, within the context of this experiment. No one in the eldest age group used P1 or P3 (Table 4.9), but P1 and P3 were used by 6 of the 8 participants in the youngest age group. P2 was the most commonly used explanation, with 32% of all trials explained by P2 (Table 4.8). P2 was also used by the greatest number of participants: 19 of the 24 participants (79.2%) provided P2 at least once. P2 was commonly used for explaining why asymmetrical blocks cannot be balanced.

Let us turn back to the interesting results of the 3 participants in the group of 8 to 9 year olds (CM1, CM3, and CM4), who successfully balanced all of the asymmetrical blocks in both Phase One and Phase Two, but predicted that all of the asymmetrical blocks could not be balanced. In the Prediction Task, when they were asked why these blocks could not be balanced, all 3 participants' responses were explanation P2.

E: Hum... Why can't this pile of blocks be balanced?

CM1: Because... This side is heavier, that side is not heavy.

(CM1, Prediction Task)

E: You think these blocks cannot be balanced on that wooden board; they will fall, right?

CM3: Because this side is lighter, that side is heavier.

(CM3, Prediction Task)

E: So this pile of blocks, if put on that wooden board, they must fall and cannot be balanced?

(CM4 nodded)

E: Why? Why can't they be balanced?

CM4: Because the two sides... two sides weigh differently.

(CM4, Prediction Task)

However, in both test Phase One and Phase Two, all three participants successfully balanced all of the asymmetrical blocks, and they provided explanation type T for their success in both Phase One and Two. Other than these three participants, the two other participants, (CF1 and CM2), who had made similar contradictions in the Prediction Task also used P2 to explain why they thought that the asymmetrical blocks could not be balanced in the Prediction Task. Even though the 6 participants had previously used explanation T and had successfully balanced asymmetrical blocks before, they still used P2 to explain why the asymmetrical blocks could not be balanced in the Prediction Task.

4.4 Comparison of behavioural performance and performance that demonstrates explicit understanding

In this section, after examining the behavioural performance and performance that demonstrates explicit understanding, the data obtained from these performances will be compared to obtain a more comprehensive picture of development in the block-balancing task.

4.4.1 Geometric-centre theory

Whether the geometric-centre theory exists, and the whether geometric-centre theory exists in implicit form, has important theoretical implications for the RR model. Behavioural performance, BGeo, revealed whether participants had persistent central placement, and explicit verbal explanations, type G, revealed whether the participants had awareness of their own geometric-centre theory. Four time points were set in order to capture changes as they occurred during the experiment. For more information about the time points and the classification labels used in Figure

14.5, please refer to section 3.1.5.3.

	Age group 4-5							
	AF1	AF2	AF3	AF4	AM1	AM2	AM3	AM4
T1	IG	N	IG	N	N	N	IG	IG
T2	**≥3**	-	**≥3**	-	-	-	**≥3**	**≥3**
T3	IG	N	IG	IG	IG	N	IG	**EG**
T4	IG	N	IG	IG	**EG**	N	IG	**EG**
	Age group 6-7							
	BF1	BF2	BF3	BF4	BM1	BM2	BM3	BM4
T1	N	IG	IG	IG	IG	IG	N	N
T2	-	**≥3**	**≥3**	**≥3**	-	**≥3**	-	-
T3	N	IG	IG	IG	IG	IG	N	N
T4	N	IG	IG	**EG**	IG	IG	N	N
	Age group 8-9							
	CF1	CF2	CF3	CF4	CM1	CM2	CM3	CM4
T1	N	**EG**	N	N	N	N	N	N
T2	-	-	-	-	**≥3**	-	**≥3**	**≥3**
T3	N	N	N	N	N	N	N	N
T4	N	N	N	N	N	N	N	N

Table 4.11 Individual development over time

*Note: "**T1**" = Phase One, "**T2**" = Prediction Task, "**T3**" = Phase Two, and "**T4**" = questioning about the initial placement position and its relation with block-balancing.*

*For T1, T2 and T4: "**IG**" = Implicit geometric-centre theory, "**EG**" = Explicit geometric-centre theory, "**E Only**" = Explicit geometric-centre theory only, and "**N**" = No geometric-centre theory found.*

*For T3: "**≥3**" = the participant predicted that 3 or more asymmetrical blocks could not be balanced, and "**-**" = the previous condition was not met.*

Shaded cells indicate the influence of geometric-centre theory at that particular time point.

4.4.1.1 Implicit geometric-centre theory and explicit geometric-centre theory

Figure 4.15 shows that 6 of the 8 participants in the group of 4 to 5 year olds, and 5 of the 8 participants in the group of 6 to 7 year olds were classified as IG for at least one time point. This result supports the hypothesis that implicit geometric-centre theory exists in younger children.

Of the 4 participants who were labelled as EG for at least one time point, three had received an IG label in the previous time point. The remaining participant who had EG at T1 belonged to the eldest age group. This suggests the possibility that the developmental order is that IG precedes EG, but the number of cases is too small to provide solid support.

In the youngest age group, 2 participants (AF4 and AM1) did not possess geometric-centre theory in T1 and T2, but developed BGeo in T3.

4.4.1.2 Implicit geometric-centre theory: level-I or E1?

After proving the existence of implicit geometric-centre theory (IG), further evidence is needed to decide whether IG is represented by level-I representation or E1 representation. According to Karmiloff-Smith (1992), neither level-I nor E1 representation can be accessed consciously and reported verbally. Level-I representation is bracketed, inflexible, and individual components cannot be singled out to share with other process. No intradomain and interdomain link is possible. Knowledge cannot be transferred and applied to other task. On the other hand, E1 representation is not bracketed, so it is available for potential intradomain and interdomain representation. E1 can be generated by abstraction from level-I representation.

If IG is a level-I representation, its effect should only be limited to T1 and T3, in which the behavioural skill can be applied. The Prediction Task should not be affected because no block-balancing action is involved in this task. If participants exhibiting BGeo also predict that the asymmetrical blocks cannot be balanced, then their IG should be represented at E1 level.

Of the 11 participants who had the IG label, 2 first had IG in T3, which means that their IG emerged after completion of the Prediction Task, so these 2 participants were excluded from the analysis. Among the 9 participants who had IG before T2, 8 (89%) of them predicted at least three times that asymmetrical blocks could not be balanced in the Prediction Task. The only participant who did not make this prediction was BM1. In the free play period, which was a section held just before the Prediction Task, the participant succeeded in balancing three of the asymmetrical blocks, so he predicted that those blocks could be balanced. But in T4, he failed to repeat the success in balancing the asymmetrical blocks and again claimed those blocks could not be balanced. Observing from the video recording, in the free play period, BM1's placements of the blocks were performed in a very casual manner, he approached the block-balancing task much more seriously in Phase One and Phase Two. Karmiloff-Smith (1992) pointed out that children possessing implicit geometric-centre theory still possessed the ability to balance asymmetrical blocks, because the level-I representation sustained this type of function still existed. When the explicit goal was house building, these children could balance asymmetrical blocks. But when the explicit goal was to balance blocks, the children called on explicit representation and this resulted in exclusive failure in balancing asymmetrical blocks. BM1's success in free play and failure in Phase One and Phase Two may also be explained in the same way. When he approached the task seriously, explicit knowledge was called upon to make the inference, and that was why he failed in Phases One and Two. When he approached the task casually, he might not have given the task so much thought and only used level-I representation to complete the task.

The implicit geometric-centre theory did not only affect the behavioural result in test Phases One and Two, but it also exerted influence on conscious inference in the Prediction Task. Given that the implicit geometric-centre theory could be used outside the normal input-output relation, it cannot be represented at level-I. However, the implicit geometric-centre theory could not be verbally reported, so it could not be represented at E2/3. Therefore, this implicit geometric-centre theory should be represented at E1.

4.4.2 Development from top-down and bottom-up

There are two possible directions of development along the explicit-implicit dimension: top-down and bottom-up. The top-down direction of development is the process of knowledge represented in the explicit level guiding the development of the implicit level. The bottom-up direction of development involves turning knowledge that is embedded in implicit level into explicit knowledge. This section investigates whether the top-down or the bottom-up direction of development occurred in the block-balancing task.

4.4.2.1 Relationship between success in balancing individual blocks and the correct explanation

If the developmental direction between behavioural success in balancing individual blocks and the correct explanations is top-down, in the first stage, a correct explanation should exist without behavioural success, and in the next stage, both correct explanation and behavioural success can be found. Explicit knowledge is the precondition of behavioural success in the top-down direction of development, that is, if there is no correct explicit knowledge, there will be no behavioural success. Therefore, a stage (or trial) in which behavioural success exists without a correct explanation should not exist. Therefore, P (Behavioural success | Explanation Incorrect) should be 0. In this experiment, explanation types T and B are correct, and type G is correct for the symmetrical blocks. Explanation type N (No explanation), P1, P2, P3, O (Others), and explanation G for the asymmetrical blocks are regarded as incorrect explanations. The total number of trials that have an incorrect explanation is 248. Among these 248 trials, block-balancing success is found in 131 trials, so P (Behavioural success | Explanation incorrect) = 52.8%.

Also, if there is no correct explicit knowledge, there will be no behavioural success, then P (Explanation incorrect | Behavioural success) should also be 0. In the experiment, there were 267 trials in which blocks were successfully balanced. Among these trials, incorrect explanations were found in 131 trials, so P (Explanation incorrect | Behavioural success) = 49.1%.

Based on the above statistics, it seems unlikely that the behavioural success of this block-balancing task is a result of top-down development.

If the developmental direction is bottom up, in the first stage, behavioural success should exist without correct explanation. In the next stage, both behavioural success and correct explanation should be found. Behavioural success is the precondition of providing a correct explanation. Therefore, there should be no stage (or trial) in which a correct explanation coexists with behavioural failure. It was found that P (Explanation correct | Behavioural failure) = 0. Therefore, in the trials in which blocks could not be balanced, all of the explanations provided were incorrect. Also, P (Behavioural success | Explanation correct) is 100%. So, in the trials in which explanations were correct, all of the blocks were successfully balanced. These findings support the hypothesis that bottom-top development occurred in this experiment.

In the Prediction Task, 6 participants predicted that the asymmetrical blocks that they had successfully balanced in Phase One could not be balanced. They had all given explanation T in Phase One, and they succeeded in balancing the asymmetrical blocks again in Phase Two. If the development of explanation T followed the top-down direction, explicit knowledge should be the guide for development of lower levels. It is very difficult to explain why these 6 participants performed better in the task that involved behavioural implementation of explicit knowledge, but performed less well in the task that involved the use of explicit knowledge. It seems that the bottom-up direction of development is more reasonable. These participants' explicit knowledge of the naïve version of the law of torque was not consolidated well. In Phases One and Two, they could reason from their own correction behaviours to be able to give explanation T. In the Prediction Task, when no block-balancing behaviour was performed, these participants were not able to use the naïve version of law of torque to infer the answer and were influenced more by the consolidated geometric-centre theory that appeared earlier in their development.

4.4.2.2 Relationship between the behavioural pattern that agrees with the naïve version of the law of torque and the explicit verbalization of the naïve version of the law of torque

The direction of development may not be the same throughout the developmental process. It is possible for the direction of development to be bottom up in the first stage, then change to top down in the next stage. We have shown in the previous section that explanation T and other correct explanations were the results of bottom-up development. This section focuses on the naïve version of the law of torque. We investigated the developmental relation between the behavioural pattern and the explicit verbalization of the concept to see which one is the prerequisite of the other.

In the previous part, the focus was on the relation between success in balancing individual blocks and the correct explanation. Analysis was based on individual trials.

In this section, the focus is on the relationship between the behavioural pattern that agrees with the naïve version of the law of torque, and the correct explanation T. Analysis is based on the test phases. Each participant had gone through the two test phases, Phase One and Phase Two, in the experiment.

The definition and prevalence of the behavioural pattern reflecting the naïve version of the law of torque (BTorque) can be found in section 4.2.5.2. Whether each phase has at least one explanation T will also be analyzed. The distribution can be found in Table 4.12.

		Have explanation T	No explanation T
Have BTorque	7	7	0
No BTorque	41	12	29
Total		19	29

Table 4.12 Distribution of BTorque and explanation T across phases

If explicit knowledge is the precondition of a behavioural pattern, then P (BTorque | No explanation T) and P (No explanation T | BTorque) should both be 0. Referring to Table 4.12, P (BTorque | No explanation T) = 0 / 29 = 0. P (No explanation T | BTorque) = 0 / 7 = 0. This shows that the verbalization of explanation type T occurred before the behavioural pattern, so this stage of development seems to follow a top-down direction.

The developmental path can be summarized as follows:

Behavioural success of individual blocks ⇨ Occurrence of explanation involving the naïve version of the law of torque ⇨ Behavioural pattern reflecting the naïve version of the law of torque

4.5 Summary

4.5.1 List of important findings related to the RR model

Findings related to the explicit-implicit dimension:

1) The performance difference between block balancing and the Prediction Task was observed in this block-balancing experiment. If both kinds of performance were

sustained by the same type of representation, then their pace of development and the factors that affect their development should be the same across age groups. However, it was found that the significant factors affecting the behavioural measures and the Prediction Task were different. For all of the behavioural measures, the main effect of age was significant, but this was not the case for the Prediction Task. The performance for the asymmetrical blocks in the Prediction Task was very similar for all three age groups. It seems more reasonable to take the position that the knowledge that sustained the behavioural performance and the knowledge that led to the prediction performance were represented at different levels rather than at the same level.

2) Evidence for supporting the existence of implicit geometric-centre theory was found.

3) The implicit geometric-centre theory was not represented at level-I representation but was represented at E1 representation, because it could exert influence outside the normal input-output relation, which influenced the results of the Prediction Task. This finding provides empirical support for the existence of the level-E1 representation, which is a unique contribution of the RR model that breaks the implicit/explicit dichotomy.

Findings related to development direction (top-down/bottom-up development):

1) It was found that explicit correct knowledge was not the precondition of behavioural success in block balancing. On the other hand, in the trials in which blocks could not be balanced, all of the explanations provided were incorrect. Among the trials in which the explanations were correct, all of the blocks were successfully balanced. It seems that the implicit knowledge about block balancing was acquired before the explicit knowledge. If that was the case, the direction of development is bottom up.

2) The developmental path of the naïve version of the law of torque was observed in this study and can be summarized from the results. If a participant first succeed in balancing some symmetrical and asymmetrical block(s), this might lead to the discovery of the naïve version of the law of torque. After this discovery, the behavioural pattern reflecting this knowledge might appear. Put a different way, bottom-up development was involved in the discovery of the explicit concept, and top-down development was involved in the occurrence of the related behavioural pattern. This result suggests that, in the developmental process of a concept, top-down and bottom-up development can happen for the same concept in different stages.

3) Six participants predicted that some or all of the asymmetrical blocks could not be

balanced, and they explained their predictions using P2. However, the blocks were successfully balanced by these participants in Phases One and Two, and they provided explanation type T in both phases. This result contradicted the thesis of top-down development, because the performance of the task involving application of explicit understanding was worse than the performance of the task requiring behavioural implementation of explicit knowledge. Therefore, the bottom-up direction of development appears to be more reasonable.

Findings related to the U-shaped performance curve:

1) No U-shaped performance curve was found for behavioural performance across age groups. According to the RR model, behavioural performance may decline while internal representation is developing to a higher level. Karmiloff-Smith (1992) said that this performance decline does not necessarily occur. Therefore, even though the U-shaped curve was not found in this experiment, this does not threaten the plausibility of the RR model.

4.5.2 List of important findings related to the age-related development of the block-balancing task

1. The behavioural performance of the group of 8 to 9 year olds was significantly

more successful in balancing asymmetrical blocks, and had significantly less initial middle placement than the groups of 4 to 5 year olds and 6 to 7 year olds. The persistent central-area placement throughout the trial of the group of 8 to 9 year olds was only significantly different from the group of 4 to 5 year olds.

2. The behavioural pattern of geometric-centre theory was concentrated in the groups of 4 to 5 year olds and 6 to 7 year olds, and was rarely found in the group of 8 to 9 year olds.

3. The behavioural pattern that reflected the naïve version of the law of torque was concentrated in the group of 8 to 9 year olds, being rarely found in the group of 6 to 7 year olds, and absent in the group of 4 to 5 year olds.

4. The performance for asymmetrical blocks in the Prediction Task was similar for all of the three age groups.

5. The explanation of the naïve version of the law of torque first occurred in the group of 6 to 7 year olds, and was given by all of the participants of the group of 8 to 9 year olds.

6. Two incorrect explanations, P1 and P3, were not found in the group of 8 to 9 year olds.

7. Implicit geometric-centre theory was commonly found among children in the groups of 4 to 5 year olds and 6 to 7 year olds, but was absent in the group of 8 to 9

year olds. The only participant who exhibited BGeo in the group of 8 to 9 year olds was aware of her geometric-centre theory.

4.5.3 Responding to key research questions in a nutshell

To summarize, the findings of this block-balancing task provided support for the RR model. Though the U-shaped performance curve was not found, this did not pose any challenge to the RR model, because Karmiloff-Smith (1992) had already said that it was not essential for the U-shaped performance to occur. More importantly, the differences between implicit and explicit representations and the bottom-up direction of development can be found in the current experiment. Evidence for the existence of the E1 level was also found. The results of experiment one provided empirical support for the RR model.

CHAPTER FIVE: RESULTS AND DISCUSSION OF STUDY TWO

5.1 Outline

The performance of the probability task will be reported from two directions: behavioural performance and performance that demonstrates explicit understanding.

In this probability task, three issues associated with behavioural performance were investigated: overall accuracy, difference between pretest and posttest performance, and behavioural performance of the half rule.

To study the performance that demonstrates explicit understanding, explanations provided by participants were classified and studied carefully, in order to understand what children explicitly knew about the task.

Behavioural performance and explicit understanding of the participants will be compared to see whether the developmental pattern of the two is different, and whether the development of probability concepts was from the bottom-up direction.

5.1.1 Notes on participants' background knowledge

Before conducting the experiment, the researcher had consulted the head teacher of the primary school about whether the participants had received any instruction about fractions. The teacher said this topic would be taught to primary 3 students, and the period assigned for teaching this topic was scheduled to be after the empirical study. In other words, none of the participants in the current study had received any instruction about fractions at school.

5.2 Behavioural performance

5.2.1 Overall accuracy score

The accuracy score of each question was obtained by calculating the distance between the response and the correct answer. The correct response scored zero. In other words, a lower accuracy score meant better performance.

The overall performance, (that is, performance including all test phases: pretest, training, and posttest,) was analyzed using ANOVA. The between-subjects factors were age group (Age groups 4-5, 6-7, and 8-9) and gender (Male and Female), and the within-subjects factor was the question number (a total of 18 questions in the three test phases).

The main effect of age group was significant, $F(2, 18) = 23.220$, $p < .05$, Partial

Eta Squared = .721. The main effect of gender, however, was not significant, F(1, 18) = .002, p> .05. The effect of question number was significant, F(17, 306) = 3.392, p < .01, Partial Eta Squared = .159. The only significant interaction was between question number and age group, F(34, 306) = 1.910, p < .01, Partial Eta Squared = .175.

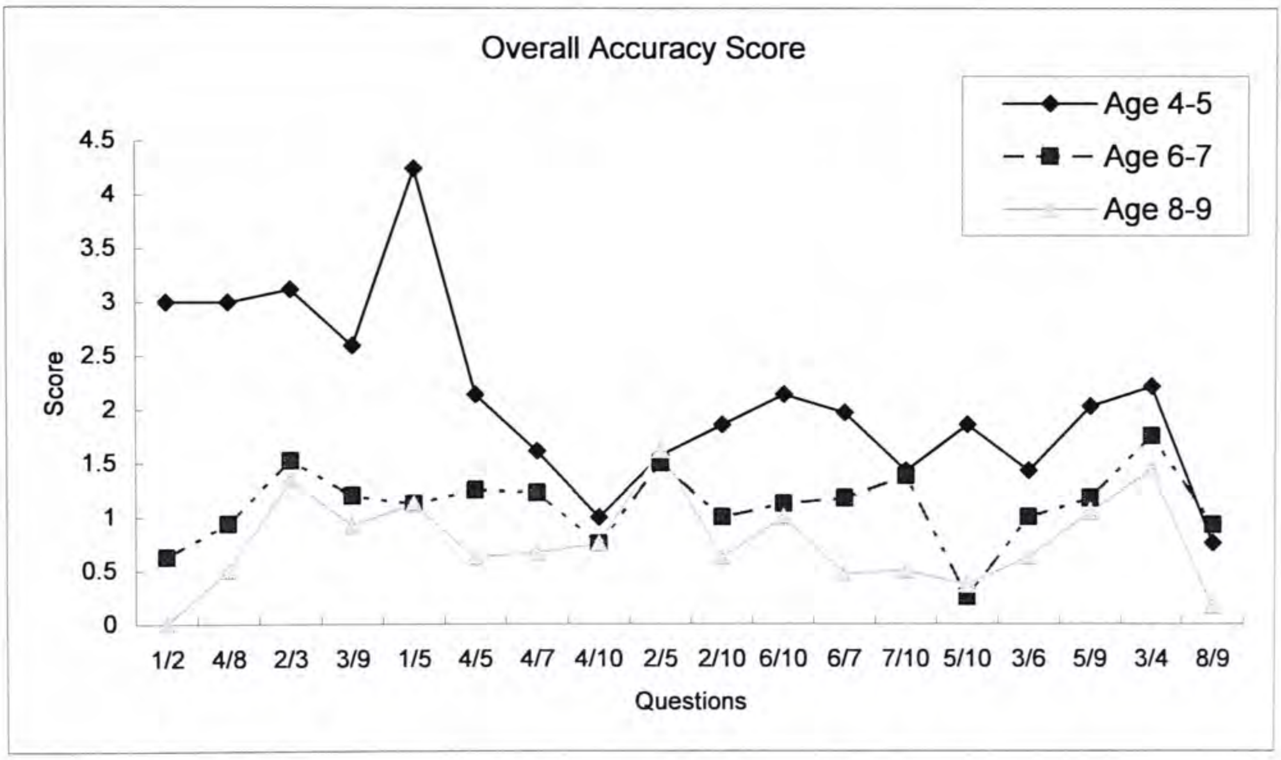

Fig. 5.1 Overall accuracy score

No U-shaped performance curve was found across the age groups. The performance improved as age increased. According to Karmiloff-Smith (1992), the reason for the U-shaped performance curve is that after reaching behavioural mastery, the implicit knowledge embedded in level-I may become available to consciousness. This awareness of knowledge may lead to overemphasis on that knowledge, resulting in a decline in performance. In the current study, though the U-shaped performance curve was absent, there was evidence to suggest that there may be improvement in awareness of previously implicit knowledge. This will be further discussed in the section related to bottom-up development.

Age 4-5		Age 6-7		Age 8-9	
Worst	Best	Worst	Best	Worst	Best
1/2 (4.25)	8/9 (0.76)	3/4 (1.75)	5/10 (0.25)	2/5 (1.625)	1/2 (0)
Pretest Q1	Posttest Q5	Posttest Q4	Posttest Q1	Training Q4	Pretest Q1

Table 5.1 Questions that had maximum or minimum accuracy score in each age group (accuracy score in brackets)

The questions that were answered best and worst by each age group are shown in Table 5.1. It should be noted that, since the primary aim of this study was not to identify which fraction participants would perform best or worst, the presentation order of the questions was fixed, not randomized among the participants. Therefore, it was difficult to rule out the practice effect as a whole (in fact, the experiment was designed to ascertain whether practice does have an effect). The performance difference found in the group of 4 to 5 year olds might be due to practice, because the best question was the last question of the whole experiment and the worst question was the first question of the whole experiment. However, it seems that the practice effect cannot completely explain the result in Table 5.1, because the presentation order was the same among the age groups, but the best and worst questions differed among the age groups. The best question for the group of 8 to 9 year olds was the first question of the experiment, the worst question for the 6 to 7 year olds was the second-to-last question of the whole experiment. These two results could not be explained by the practice effect. In an experiment that used the proportional matching paradigm, Singer-Freeman and Goswami (2001) also found that among participants of 3 to 4 years old, their performance for problems involving halves and three-quarters were better than for problems involving quarters. Different fractions have a different level of difficulty, and this level of difficulty interact with age.

5.2.2 Pretest and posttest score

In the training phase of the experiment, correct answers were provided to the participants after the participants had given their own answer. Other than the correct answer, the experimenter did not provide any other information or instruction about how the computation should be carried out. Therefore, if there was improvement in the posttest, this was not the result of explicit knowledge taught to the participants, and should have been the result of experience gained, particularly from the interaction with feedback, in the training phase.

ANOVA of accuracy score was carried out. The between-subjects factors were age group (Age groups 4-5, 6-7, and 8-9) and gender (Male and Female), and the within-subjects factors were phase (pretest and posttest) and question type (questions 1-5 of each test phase). The questions in pretest and posttest were designed to be parallel: please refer to the method section 3.2.3 for a detailed explanation.

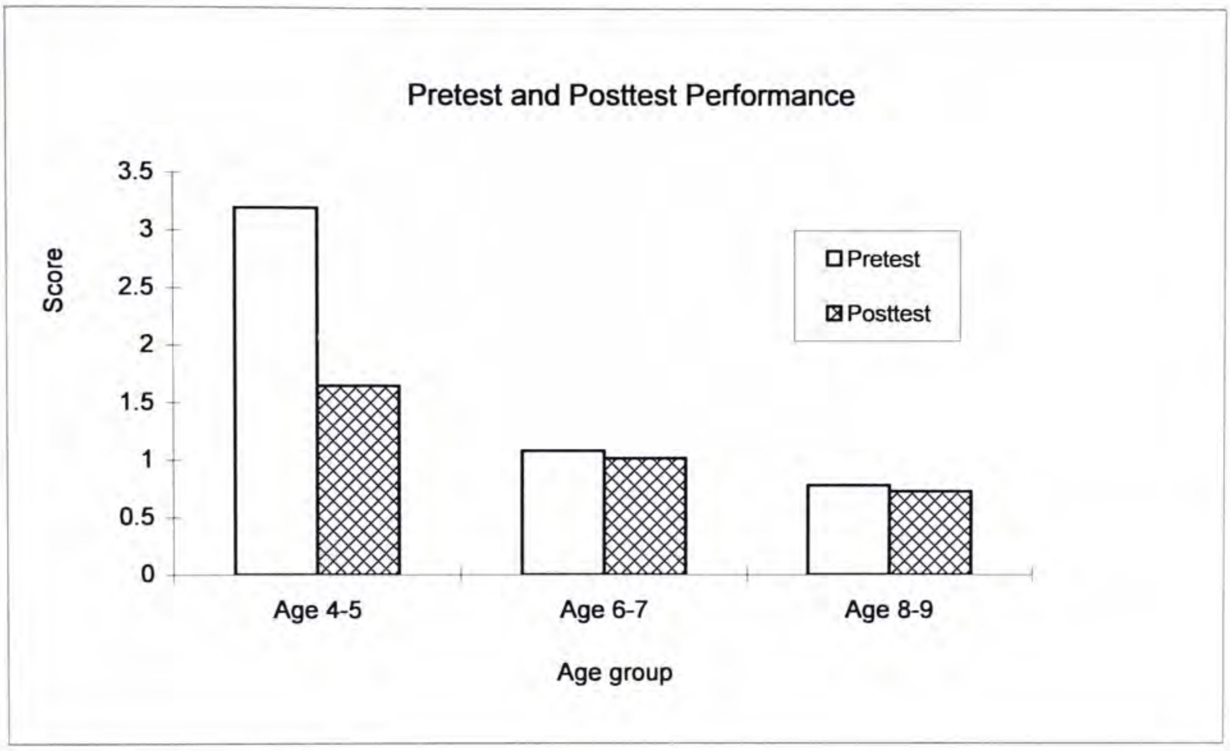

Fig. 5.2 Difference between pretest and posttest performance

The effect of phase was significant, $F(1, 18) = 8.891$, $p < .05$, Partial Eta Squared = .331. The effect of question type was also significant, $F(4, 72) = 4.29$, $p < .05$, Partial Eta Squared = .193. The main effect of the between-subjects factor age group was significant, $F(2, 18) = 26.770$, $p < .05$, Partial Eta Squared = .748. The main effect of gender, however, was not significant, $F(1, 18) = .009$, $p > .05$.

Only the following two interactions are significant: the interaction effect of phase and question type was significant, $F(4, 72) = 4.103$, $p < .05$, Partial Eta

Squared = .186, and the interaction between phase and age group was significant, $F(2, 18) = 7.173$, $p < .05$, Partial Eta Squared = .444. Using Bonferroni adjustment to control the familywise alpha at .05 level, pairwise comparisons indicated that only the group of 4 to 5 year olds shows a significant difference between the pretest and the posttest (Mean difference: pretest – post test = 1.545, SD = .321, $p < .01$).

It can be seen that the participants in the group of 4 to 5 year olds performed significantly better in the posttest. One of the possible reasons was that this youngest group of participants might not have fully understood the task in the first few trials, resulting in poorer performance in the pretest. In the posttest phase, their understanding was improved because of the experience gained from the pretest and training phases. It should be noted that the experimenter took extra care to ensure that these participants understood. The experimenter carefully inspected the explanations provided by these participants, particularly in the first few trials. If there was any sign of confusion or misunderstanding, the experimenter explained the whole task again.

Pretest		Training				Posttest	
Q no	Score	Q no	Score	Q no	Score	Q no	Score
1	3.00	1	2.14	5	1.86	1	1.86
2	3.00	2	1.61	6	2.14	2	1.43
3	3.13	3	1.00	7	1.97	3	2.03
4	2.60	4	1.57	8	1.43	4	2.21
5	4.25					5	0.76

Table 5.2 Accuracy Score of the group of 4 to 5 year olds

Table 5.2 lists the accuracy score of each question for the group of 4 to 5 year olds. If the poor performance in pretest was due to incomplete understanding of the task instructions, and the experience gained from trials led to behavioural improvement by improving the understanding of the instructions, then the improvement of accuracy score should take place gradually across the trials. However, it seems that Table 5.2 does not exhibit this pattern, because in the pretest, a general trend of improvement across the five trials cannot be observed. The performance of question 5 was even worse than the performance of question 1. In the training phase, the accuracy score remained about 1-2. The scores were much smaller when compared with the pretest. As mentioned before, correct answers were provided in the training phase. It seems more likely that the participants were able to make use of the correct answer to fine-tune their estimation. If this was the case, then the improvement found in the posttest might be caused by the fine tuned estimation ability.

Among the three age groups, only the youngest age group was able to make significant improvement in the posttest phase. Why did the two older age groups fail to benefit from the training phase? One of the possible explanations is the difference in initial performance. In figure 5.2, it can be seen that the initial performance of the youngest age group was much worse than the two older age groups. Therefore, the youngest age group had more room for improvement. The two older age groups may have reached the ceiling performance for intuitive estimation. To achieve further improvement, correct explicit knowledge must be acquired. The training phase only provided correct answers: no direct instruction on correct computational procedure was provided. The correct answers were more helpful in fine-tuning the intuitive estimation, but they did not provide any direct information about the correct computational procedure. Therefore, only the youngest age group could benefit from the training.

More details about this improvement will be discussed later in this chapter.

5.2.3 Behavioural adherence of the half rule

The behavioural half rule score (BHalf score) of each participant was obtained by counting the number of trials that adhered to the half rule in all three phases. For a detailed explanation, please refer to section 3.2.5.1. ANOVA of BHalf score was implemented. The between-subjects factors were age group (Age groups 4-5, 6-7, and 8-9) and gender (Male and Female).

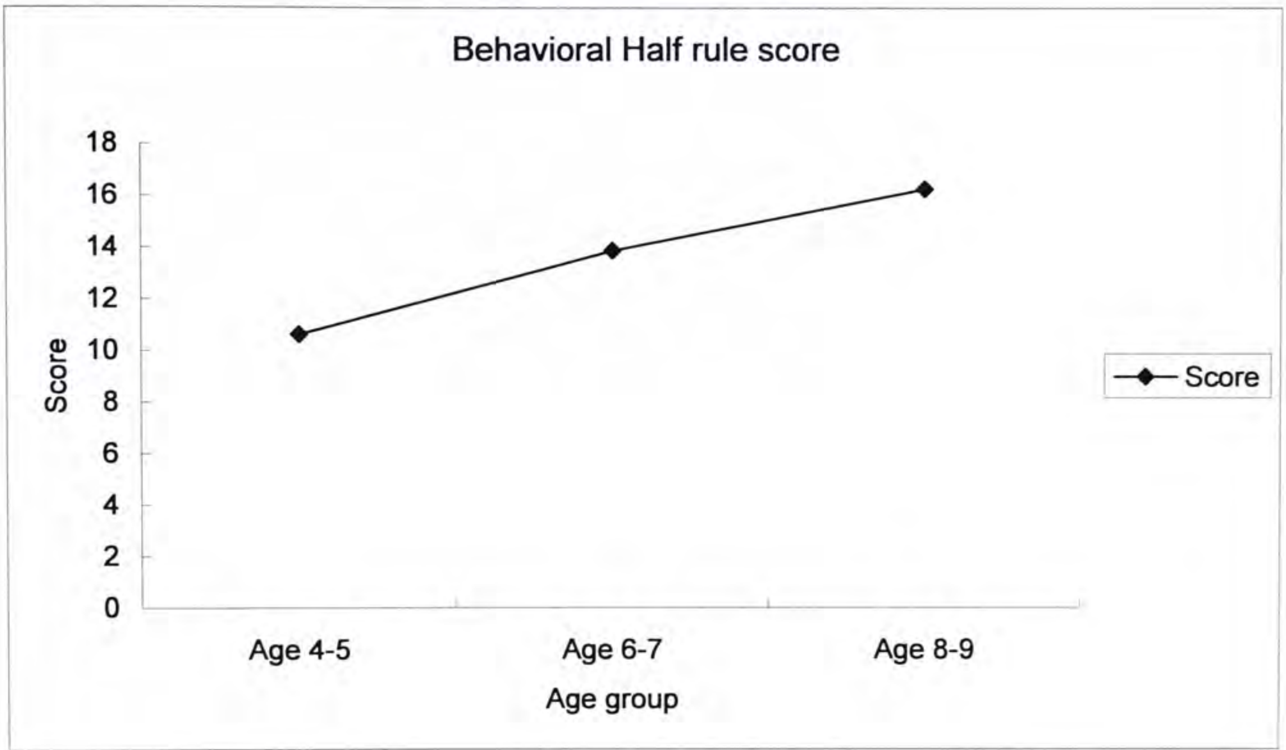

Fig. 5.3 Behavioural Half rule score

The effect of age group was significant, $F(2, 18) = 16.763$, $p < .01$, Partial Eta Squared = .651. The effect of gender, however, was not significant, $F(1, 18) = .175$, $p > .05$. The interaction between age group and gender was not significant either, $F(2, 18) = .405$, $p > .05$. The post hoc tests using the Tukey HSD (honestly significant difference) procedure indicated that the performance of the group of 4 to 5 year olds was significantly different from both the 6- to 7-year-old and 8- to 9-year-old groups. However, the difference between the group of 6 to 7 year olds and the group of 8 to 9 year olds was not significant.

We will revisit the BHalf score later in this chapter. The development of BHalf score will be compared with explicit performance of the half rule, to see whether the two improved at the same rate among different age groups.

5.3 Performance that demonstrates explicit understanding

5.3.1 Verbal explanations

In each trial, after participants had provided their answers, the experimenter asked them to explain their answers. The explanations were classified, based on the computational methods used, into one of seven different categories: 1) Implicit computation (I); 2) Last question (L); 3) Comparison (C); 4) Plus minus (P); 5) Half rule (H); 6) Division/Fraction (F); and 7) Others—Idiosyncratic (O). For more information about the classification requirements of these categories, please refer to section 3.2.5.2.

The distribution of explanation types are shown in Figure 5.4 and Tables 5.3 to 5.5. As reported in section 3.2.5.2, the interrater agreement on the explanation types is substantial.

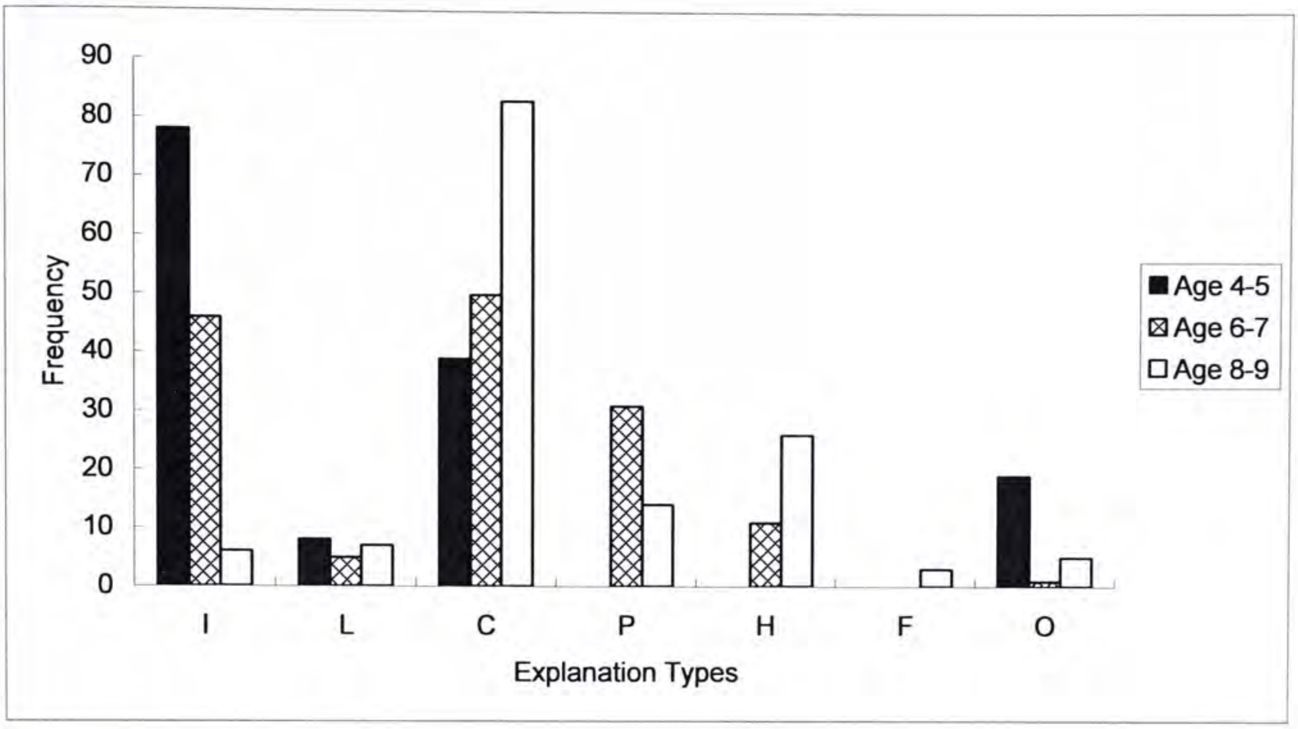

Fig. 5.4 Distribution of explanation types among trials (in terms of frequency)

	Age 4-5		Age 6-7		Age 8-9		Total	
I	78	(54.2%)	46	(32.9%)	6	(4.2%)	130	(30.1%)
L	8	(5.6%)	5	(3.5%)	7	(4.9%)	20	(4.6%)
C	39	(27.1%)	50	(34.7%)	83	(57.6%)	172	(39.8%)
P	0	(0%)	31	(21.5%)	14	(9.7%)	45	(10.4%)
H	0	(0%)	11	(7.6%)	26	(18.1%)	37	(8.6%)
F	0	(0%)	0	(0%)	3	(2.1%)	3	(0.7%)
O	19	(13.2%)	1	(0.7%)	5	(3.5%)	25	(5.8%)

Table 5.3 Distribution of explanation types among trials (in terms of percentage)

	Age 4-5	Age 6-7	Age 8-9
I	8	8	4
L	2	1	5
C	7	6	8
P	0	3	4
H	0	5	8
F	0	0	2
O	6	1	4

Table 5.4 Distribution of explanation types among participants (8 participants in each age group)

	Age 4-5	Age 6-7	Age 8-9	Across age group
I	2.159	0.963	0.650	1.666
L	1.875	1.060	1.043	1.380
C	1.967	1.050	0.999	1.233
P	-	1.684	0.629	1.356
H	-	0.145	0.158	0.154
F	-	-	0.000	0.000
O	2.311	3.400	0.720	2.036

Table 5.5 Accuracy score of each type of explanation

Among the seven types of explanations, only type P (Plus minus) and F (Division/Friction) consisted of an exact computational procedure that could be repeated. Half rule (H) guaranteed the same answer when the number of both types of grasses was the same. For the other situations, the half rule only provided a range for the correct answer.

5.3.1.1 Explanation type F (Division/Fraction)

An explanation is classified as type F (Division/Fraction) if participants used the concept of division or fractions to explain how they arrived at their answers. Here is an example of type F:

(5 green grasses and 5 brown grasses, correct answer: 5)

E: 5 points, why is it 5 points?

CF3: Because... if all of them are green then it is 10 points. But now it is equally divided, half is green, half is brown, so 10 divided by 2 is 5 points.

(CF3, Posttest Question 1)

The accuracy score of all of the responses that were explained by type F explanation were 0 (Table 5.5), that is, all responses were completely correct. This type of explanation was only found in the eldest age group of 8 to 9 year olds, and it was only used in three trials (Table 5.3) by two participants (Table 5.4). All of these trials had equal numbers of green grasses and brown grasses. Although this explanation could be used for all of the questions in the experiment, neither participant completely shifted to explanation F after its first occurrence. Participant CF1 used this explanation for the first trial of the experiment, but after that she never used it again, and instead she offered explanation types L, C, and H in later trials. Participant CF3 first used explanation type F in the first question of the posttest, then she used it in the second question, but after that she fell back to using explanations H and P.

Participant CF1 used the concept of fractions to explain her answer. As mentioned before, the school had not taught this topic yet, so the experimenter was curious about the source of her concept:

E: I see, the chance is 5/10. Have you learnt about fractions? (CF1 shook her head) But why do you know things like "something over ten"?

CF1: Because my mother has taught me.

(CF1, Pretest Question 1)

This is an example of children using explicit knowledge in problem solving, and the direction of development reflected in this trial is top-down development. However, it should also be noted that the area of application for this concept was very narrow for CF1. She only used explanation F for this question, and fell back to explanations L, C, and H after the first trial. It suggests that the more advanced knowledge coexisted with less advanced knowledge in CF1's mind, and the more advanced knowledge was mastered less well than the less advanced knowledge, because the application area of the advanced knowledge was narrow.

5.3.1.2 Explanation type H (Half rule)

The type H explanation (Half rule) requires participants to link the given situation of "the two types of grasses are equal in number" to "5 points on the Answer Board"—the middle point along the 10 point scale. Explanation H and explanation C (comparison) were very similar in the sense that both types of explanations involved comparison between the quantities of two types of grasses. The major difference between type H and type C was the "anchoring" of the "half" in the grasses situation with the "half" on the 10 point scale, and realizing that when the number of green grasses was more, the answer must be greater than 5, and when the number of brown grasses was more, the answer must be smaller than 5. Here is an example of explanation type H:

(6 green grasses, 4 brown grasses, correct answer: 6)

E: Why is it 7 points?

CF3: Because... there are 4 distasteful grasses, delicious grass... there are 6. Therefore, the chance for it to get green grass is bigger. So, if half and half should be 5 points, now it has 2 more, so it is 7 points.

(CF3, Training Questions 6)

The accuracy score of explanation H was 0.154 (see Table 5.5), which was much lower than explanation I (1.666), L (1.380), C (1.233), and P (1.356). Explanation H was the second most accurate type of explanation, being only less accurate than explanation T.

No participant in the group of 4 to 5 year olds was capable of providing explanation H. This agrees with Spinillo and Bryant's (1991) finding that 4- and 5 year olds were not capable of using the half boundary to make judgments about proportion. In the group of 6 to 7 year olds, five of the 8 participants provided explanation H, but this explanation only appeared in 7.6% of the trials (Table 5.3). All of the participants in the group of 8 to 9 year olds gave explanation H, and it was the second most popular explanation, occurring in 18.1% of the trials (Table 5.3). This agrees with Spinillo and Bryant's (1999) finding that 6 to 8 year olds were able to make use of the half boundary for making comparisons between the ratios of discontinuous quantities. The current research extended the application area of the half boundary. In Spinillo and Bryant's (1991, 1999) experiment, one target ratio was presented (represented by white and blue areas of different sizes within a rectangle, or by different size slices of a pizza), participants were required to choose from two ratios that matched the target. In the current study, the requirement of the question was to convert discrete numbers of grasses into a 10-point scale.

	Question for which H first occurs		Occurrence order of other explanations after H (Counting from the trial in which H first occurred)				
Subject			1	2	3	4	5
BF3	3/6	(Posttest Q2)	C				
BF4	5/10	(Posttest Q1)	C	I			
BM1	1/2	(Pretest Q1)	P	I			
BM3	5/10	(Posttest Q1)	C				
BM4	4/8	(Pretest Q2)	I	C			
CF1	4/8	(Pretest Q2)	C				
CF2	1/2	(Pretest Q1)	C	P	L		
CF3	5/10	(Posttest Q1)	C	O	L	P	T
CF4	1/2	(Pretest Q1)	C				
CM1	1/2	(Pretest Q1)	C	P			
CM2	1/2	(Pretest Q1)	L	C	I		
CM3	1/2	(Pretest Q1)	I	C	P	O	
CM4	1/2	(Pretest Q1)	L	O	C		

Table 5.6 The trials in which explanation H first occurred and the types of explanation found after the first occurrence of explanation H

Like explanation F, the trials in which explanation H first occurred were trials in which the numbers of green grasses and brown grasses were the same (see Table 5.6). This type of question was the easiest for the children. The finding that more advanced explanations (F and H) all first occurred in easier questions is very different from the related finding in the block-balancing task. In the block-balancing task, the more advanced explanations (Explanation T, the naïve version of the law of torque) all first occurred in trials of asymmetrical blocks, which were the more difficult tasks. It seems that whether the more advanced explanation first occurs in more difficult trials or easier trials depends on the nature of the task.

Explanation H could also be applied in all trials, but after the first occurrence of explanation H, no participant completely shifted to explanation H, and all of them used less advanced explanations (e.g., I, L, C, and P) in later trials. This finding agrees with the finding of the block-balancing task that lower explanations were only phased out gradually.

5.3.1.3 Explanation type P (Plus minus)

Participants who used explanation type P assigned a particular score for each green or brown grass, and their result was obtained by adding or subtracting these scores. There were two main subtypes of explanation P (where GG = Green grass and BG = Brown grass):

Subtype 1: Score of a GG **x** Number of GG – Score of a BG **x** Number of BG

Subtype 2: 10 – Score of a BG **x** Number of BG/0 + Score of a GG **x** Number of GG

If participants used subtype 1, usually they first established the ratio that one grass carried X points, then they counted the number of green grasses, and increased the number of points according to the assigned ratio at each count. Here is an example:

(4 green grasses, 6 brown grasses, correct answer: 4)

E: So, what do you think, how happy is Little Horse?

(BM1 counted the marks on the Answer Board, started from 0, counted 2 units on the board at each count. After counting to 8, BM1 started counting backward, 1 unit for each count, counted to 2.)

BM1: 2 points.

E: 2 points. Why is it 2 points?

BM1: Because I did it like the previous question, 1, 2, 3, 4 (Counting 2 units on the board as one count, reached 8), then 6 (brown) grasses, 1, 2, 3, 4, 5, 6 (Counting backward from 8, 1 unit for each count, reached 2) I did the *subtraction this way.*

(BM1, Training Question 3)

This computation procedure reflected that the participant understood that the number of both types of grass should be used to determine the probability of getting green grass. The participant had the concept of ratio, which is "1 grass, 2 points" in the example. The result was obtained by replicating; that is, adding to each set the corresponding unit for the set so that the invariant one-to-many correspondence is maintained (Nunes & Bryant, 1996, p. 145). The participant in the example carried out replication by constantly counting 2 units for each count. The ratio of how many points each grass should carry was set intuitively by the participants. None of them used division to calculate the "point per grass." When participants found that they were wrong (in the training phase), they usually thought the ratio was incorrect, and continued to use the same computation procedure in the next trial with a new ratio.

Explanation P subtype 1 was used by two participants only. One of the participants did not understand the probability meaning of "10 points": the event of "getting green grass" must happen. Once the points exceed 10, it is meaningless. For example:

(6 green grass, 4 brown grasses, correct answer: 6)

E: OK, why do you think it is 10?

BM3: More than 3 grasses (BM3 mentioned in previous trials that if the number of green grass is more than 3, then each green grass should be 3 points), 3, 6, 9, 12, 15, 18, 17, 16, 15, 14. There is no 14, so it is 10.

(BM3, Training Question 6)

BM3 did not recognize the conceptual importance of 10. He thought the answer should be 14, only there was no 14 on the Answer Board, therefore he had to opt for the closest choice: 10.

Participants who used subtype 2 of explanation P first reasoned that if all of the grasses were green, then the answer should be 10 points. If there were X brown grasses in the question, the participant then deducted the total brown grass score from 10. Some participants did the calculation the other way round. They first reasoned that if all of the grasses were brown, then it should be 0 points. Then, they added back the score of all of the green grasses. For example:

(4 green grasses, 1 brown grass, correct answer: 8)

E: Nine points isn't it? Why is it 9 points?

CM1: If there is no distasteful grass then it is 10 points, but there is one, so *deduct one point.*

(CM1, Training Question 1)

This computation procedure reflected the participants' grasp of the probability meaning of "10 points"/"0 points," and the participant was able to make use of this as an "anchor point" to carry out reasoning. However, in the trials in which explanation P subtype 2 was used, no participant used division to calculate the "point per grass." Many of them believed, like CM1 in the example, one point should be deducted for each brown grass. Some of the participants did adjust the ratio, but it was still performed intuitively, or by copying the correct ratio of the last trial. If the participants first divided 10 by the total number of grasses to get the correct "point per grass," they could get the correct answer by using this subtype 2 procedure; however, no participant using explanation P did so.

No participant in the group of 4 to 5 year olds ever provided explanation P. Of the 8 participants in the group of 6 to 7 year olds, three participants provided explanation P (Table 5.4), while two participants mainly used subtype 1 and one participant mainly used subtype 2. Explanation P is found in 21.5% of the trials performed by the group of 6 to 7 year olds (Table 5.3). Four of the 8 participants in the group of 8 to 9 year olds used explanation P, while all of them used subtype 2.

Explanation P was found in 9.7% of the trials performed by this group.

As shown in Table 5.5, the accuracy score of explanation P (1.356) is approximately the same as that of explanation L (1.380) and explanation C (1.233).

5.3.1.4 Explanation type C (Comparison)

An explanation was classified as type C (Comparison) if the participants justified their answers by comparing the number of green grasses and the number of brown grasses. The following is an example of type C explanation:

(6 green grasses, 4 brown grasses, correct answer: 6)

E: 3 points of happiness, right? Why is it 3 points?

AF2: Because 6 is more than 4.

E: 6 is more than 4... You mean there are 6 green grasses, which are more than 4 brown grasses, right?

AF2: Yes.

(AF2, Training Question 6)

In some earlier research that used the choice paradigm (Chapman, 1975; Ross & Hoemann, 1975), it was found that concrete operational children only considered one variable when they made a probability judgment. For example, there were two jars containing different combinations of marbles. If children were required to decide which jar yielded the higher probability of drawing a marble of a certain colour, they only considered the absolute number of the type of marbles that they wanted to select. They did not consider the number of the other marbles. However, the current study found otherwise. When explanation C was used, numbers of both types of grasses were considered. Explanation type C was found in all three age groups and even the youngest age group made use of this type of explanation in 27.1% of the trials.

Though participants using explanation C were aware that the number of both types of grasses had an effect on the result, the proper quantification of the probability concept had not been built up yet. Merely recognizing that "there are more green grasses" or "there are more brown grasses" could not determine the exact answer. The participant in the example correctly recognized that there green grasses were more, but the answer she gave violated the half rule. This is the difference between participants who had mastered the half rule and participants who were just capable of doing comparison: the half rule could be used to determine the range of the correct answer, but comparison itself could not suggest a range.

As shown in Table 5.3, the percentage of trials in which explanation C could be found seems to increase with age: 27.1% (age 4-5), 34.7% (age 6-7), and 57.6% (age 8-9). The researcher of this study had considered whether the participants of the group of 8 to 9 year olds in fact had explanation H in their mind, they might have other reasons for not explicitly mentioning that the answer should be "greater than 5," or "smaller than 5" (which is a classification requirement of explanation H), like this was too obvious. This possibility was also the main source of discrepancy in the explanation classifications between rater 1 (i.e., the researcher) and rater 2. Twelve explanations were classified as "Half rule" by rater 2, but classified as "Comparison" by rater 1. It was reasonable for rater 2 to make such a classification, because the explanations appeared after the first occurrence of explanation H. However, inspection of the behavioural result suggested that care should be taken, because after the first occurrence of half rule, the participant still violated the half rule in later trials. Of 13 participants who had provided explanation type H, only 3 of them completely adhered to the half rule behaviourally after the first occurrence of explanation H. Therefore, the researcher had decided to use a more conservative classification rule, by classifying the explanation as type H only when "greater than, smaller than, or equal to 5" was explicitly mentioned in the explanation.

5.3.1.5 Explanation type L (Last question) and I (Implicit)

Both explanation types L (Last question) and I (Implicit) can be found in all three age groups (see Table 5.3). The popularity of explanation I in the trials seems to decrease with age: 54.2% (age 4-5 years), 32.9% (age 6-7 years), and 4.2% (age 8-9 years). All participants of the groups of 4 to 5 year olds and 6 to 7 year olds tried this explanation, and only 4 participants in the group of 8 to 9 year olds provided this explanation. In the current study, explanation I did not only indicate a deficiency in participant knowledge, it also indicated a deficiency in metacognitive knowledge. It was expected that children of this age could not provide a completely correct explanation, which is why most of the explanation types in the current classification system were for explanation that were not completely correct. The difference between explanation I and other types of incorrect explanations (except type L) was that, in other categories, the exact computation method, or the relation suggested between the elements may be incorrect. Nevertheless, the explanations had included information that was provided along meaningful dimensions that helped to structure the problem of "how to compute probability." However, in explanation I, such meaningful dimensions were absent. In fact, no participant used explanation I throughout the experiment. Therefore, all of the participants must have some sense what kind of information was relevant to the computation of the probability problem.

Otherwise, the meaningful dimensions would not appear in the explanations that the participants provided in other trials. In the other words, although the participants made a decision about the answer, and they had some sense how the problem could be solved, they failed to recognize the information, or failed to recognize the relevance of such information, and so they failed to make use of the knowledge when they gave an explanation I. The decrease of explanation I indicates that this kind of metacognitive ability seems to increase with age.

Kuhn, Garcia-Mila, Zohar, and Andersen (1995) proposed a distinction between metacognitive knowledge and metastrategic knowledge. They explained that metacognitive knowledge involves awareness of and reflection on the content of one's thought, and reflection of the propositions that one believes to be true or chooses to consider. Metastrategic knowledge involves awareness and management of the strategies that are applied in the course of thinking and problem solving.

Explanation type L reflected metastrategic knowledge, because it reflected that the participant did not view each problem as an individual, unrelated problem. Participants using explanation L were capable of using cross-trial information to improve their understanding. Even in the youngest age group, two participants were capable of providing this kind of explanation. It seems that this metastrategic knowledge could appear quite early.

5.3.1.6 Explanation type O (Others—Idiosyncratic)

An explanation was classified as type O if a participant had mentioned a computation method or relationship between quantity of grasses, and if the response could not be classified as any of the previous categories. Within this type, the variation among explanations was quite large and to generalize the characteristics of the explanations is difficult. One common subtype in this category was the participants commented on the absolute number of green or brown grasses, like describing the absolute number as "many" or "not many," but made no comparison with the other type of grass. The percentage of trials for this explanation was 13.2% for the group of 4 to 5 year olds, 0.7% for the group of 6 to 7 year olds, and 3.5% for the group of 8 to 9 year olds.

5.3.2 Responses to the general questions after the posttest

After the completion of the posttest, the experimenter asked the participants two main questions: 1) when they tried to solve the questions, did they think about the number of green grasses, brown grasses, and total number of grasses? 2) After considering the above things, how did they know the answers?

	Age 4-5	Age 6-7	Age 8-9
Consider green grasses	6	8	8
Consider brown grasses	7	8	8
Consider total grasses	2	6	1

Table 5.7 Participants' responses to the general questions about what kinds of information they considered

All participants in the 6- to 7-year-old group and the 8- to 9-year-old group responded that they thought about both the number of green grasses and the number of brown grasses. Cross checking with their in-trial explanations, except for BM2, all of them used explanation C or P subtype 1 in the experiment, and these two explanations involved the number of the two types of grasses. The explanations provided by BM2 also involved the number of the two types of grasses. The self-reports about this question from the two age groups were accurate. In the group of 4 to 5 year olds, AF2 responded that she did not consider the number of green grasses, and AF3 responded that she considered neither the number of green grasses nor brown grasses. After checking the transcript, it was found that 12 explanations provided by AF2 involved the number of both types of grasses. Among the explanations provided by AF3, 11 explanations involved the number of one type of grass, and four explanations involved the number of both types of grasses. The self-report of this age group was less reliable. All other participants in this age group reported that they had considered both types of grasses, and this agreed with the transcript's record.

In the 4 to 5 year old group, two participants claimed that they thought about the total number of grasses, and six participants in the 6 to 7 year old group made the same claim. However, after checking with their in-trial explanations, only BM1, of the 6- to 7-year-old group, mentioned the total number of grasses in his explanation in one trial, and none of the other participants mentioned the total number of grasses in their explanations. In the 8 to 9 year old group, only one participant said he thought about this information, and this claim matched with his in-trial explanations.

Age group 4-5			Age group 6-7			Age group 8-9		
Subject	Explanation		Subject	Explanation		Subject	Explanation	
AF1	I, O, C	(O)	BF1	I, L, C	(C)	CF1	T, L, C, I, H	(C)
AF2	O, C, I	(I)	BF2	I, C	(I)	CF2	O, H, C, P, L	(C)
AF3	O,I, C, L	(I)	BF3	I, C, H	(C)	CF3	H, C, O, L, P, T	(H)
AF4	I	(I)	BF4	I, C, H	(C)	CF4	I, C, H	(H)
AM1	O, I, C, L	(I)	BM1	H, P, I	(P)	CM1	H, C, P	(H)
AM2	I, O, C	(I)	BM2	I, P, O	(I)	CM2	H, L, C, I	(H)
AM3	I, O, C	(I)	BM3	I, P, C, H	(H)	CM3	H, I, C, P, O	(C)
AM4	I, C	(I)	BM4	H, I, C	(C)	CM4	H, L, O, C	(C)

Table 5.8 Explanations provided by the participants. The abbreviations in brackets were responses provided to the question about general computational strategy. The list of explanations contains in-trial explanations provided by the participants and the order of the explanations in the lists follows the order of occurrence in the experiment.

The participants' self-reports at the end of the experiment about their general computation strategies did not fully reflect their in-trial performances. The explanations provided in response to the general questions were less advanced than the explanations that they provided in the experimental trials. In the group of 4 to 5 year olds, although 7 of the participants provided explanation C in experimental trials, none of the participants was able to tell the experimenter how they obtained the answer at the end of the experiment. In the group of 6 to 7 year olds, five participants provided explanation H in the experimental trial; however, only one reported that at the end of the experiment. One participant (BM2), who used explanation P, failed to report this procedure at the end of the experiment. In the 8 to 9 year old group, 4 of the 8 participants reported explanation H, which they had also used in experimental trials.

Explicit knowledge is the starting point of top-down development and is the guide for carrying out computation for each individual problem. The acquisition of explicit knowledge about the general principle of computation is a prerequisite to the formation of lower level representations (E1 or level-I). Therefore, if top-down development occurred, the general principle of computation should be more advanced than the in-trial explanation, which involved the application of general principles.

However, the finding of this current study suggested the opposite: the participants' explanations were less advanced in response to general questions than in the experimental trials. The alternative picture suggested by the current probability task may be explained by the possibility that when participants were faced with the question, task-specific knowledge that was not completely explicit to the participant was activated to solve the problem. This knowledge might be represented at level-I or E1 because the participants did not have full direct access to this knowledge. In the trial, the participants might have partial awareness of the activated nonexplicit knowledge, or they might apply reasoning to their own behaviour to provide an explanation. At the end of the experiment, neither of these information sources was available; therefore, the explanations provided by participants were less advanced at the end of the experiment when compared to in-trial explanations. The E2/3 representation of knowledge for this task was developed later than the non-E2/3 represented knowledge. Therefore, when the two developmental directions were compared, it seems that the bottom-up direction is the more plausible one.

5.3.3 Counter suggestions in the posttest

In the posttest, after the participants had provided answers for question 1 and 2, the experimenter gave a counter suggestion to see if the participants would change their minds. In the slide for question 1, there were 5 green grasses and 5 brown grasses. In the slide for question 2, there were 3 green grasses and 3 brown grasses. The correct answer for both of the questions was 5. If the difference of a participants' answer for questions 1 or 2 was -/+1, the experimenter reminded them that there were more green grasses in the last question, and asked why the answers to the two questions were so similar. If the difference in the participants' response was grater than 1, the experimenter reminded them that both types of grasses were equal in number in both questions, and asked why the answer was so different.

	Difference of Q1 & Q2			Responses to the counter suggestion			
	Within 1	0	Total	Change answer	Stating number of grasses equal	Others	Total
Age 4-5	2	1	3	2	1	0	3
Age 6-7	4	2	6	0	5	1	6
Age 8-9	2	6	8	0	8	0	8

Table 5.9 Number of participants whose difference in answer to Q1 and Q2 is within -/+1 and the number of responses to the counter suggestions

In the group of 4 to 5 year olds, only three participants provided similar answers for questions 1 and 2, and only one participant produced the same response for questions 1 and 2. This participant, AM1, responded with "6" for both of the questions. This answer violated the half rule. AM1 responded to the experimenter's counter suggestion by stating that the number of green grasses was equal to the number of brown grasses. The other two participants in the same age group were misled by the counter suggestion and changed their answers.

In the group of 6 to 7 year olds, 6 of the participants produced similar answers for questions 1 and 2, and 5 of them were capable of responding to counter suggestions by stating that the number of both grasses was equal. Only two participants, BM3 and BM4, provided the same value for both questions, and the value they provided was "5," this adhered to the half rule. The other four participants provided different values, and one of them, BF4, other than stating that the number of grasses was equal, also said that the number of green grasses was more in question 1, so the chance of getting green grass would be higher in question 2.

In the group of 8 to 9 year olds, all participants provided similar values for both questions, and they all responded to the experimenter's counter suggestions by indicating that both types of grasses were equal in number. However, there were still two participants who provided different values for questions 1 and 2, and both of them explained that the number of grasses in question 2 was fewer, so the answer should be smaller.

In the youngest age group, the children who were misled to change their answers all provided explanation I for questions 1 and 2. Their lack of E2/3 knowledge may be the cause of their failure to resist counter suggestions and they may have had the ability to produce nearly correct answers, but they failed to recognize that their original answers were nearly correct. In the groups of 6 to 7 year olds and 8 to 9 year olds, three of the participants believed that the absolute number of green grasses was more in question 1, so the chance of getting green should be bigger. In fact, one of the participants, CF2, explained the half rule very clearly in question 1. In question 2, she stated the half rule again, but she also stated that the absolute number of grasses had an effect on probability. She was not aware of the contradiction between the two beliefs:

CF2: Because there is equal number of delicious and distasteful grasses this time, so it is not too unhappy, and it is not too happy, so it is 5. Because, if the score is less than 5, there is less delicious grass, but it is not the case. If the score is after 5, that means 6 to 10, it is because there are less distasteful grasses, but it is not the case. So it is 5.

(CF2, Posttest Question 1)

CF2: Because there are equal number of delicious and distasteful grasses, so it... and the grasses of this time is less than last time, so its happiness level should similar to last time. It should be a little bit lower, but 3 points is too low, so I think it is 4.

(CF2, Posttest Question 2)

This example illustrated that the acquisition of the correct concept did not lead to automatic deletion of the wrong concept. It showed that a less advanced explanation was not completely wiped away when the correct explanation appeared. This also echoed another finding of both experiments in this thesis: when participants discovered more advanced or even correct explanations (Explanation T in the block-balancing task, and Explanation F and H in the probability task), they did not completely shift to the correct explanation. The lower level explanations were only phased out gradually.

	Comparing Q3 (5 green 4 brown) & Q4 (3 green 1 brown)	
	Q4 > Q3	Q3 $\leq$ Q4
Age 4-5	5	3
Age 6-7	5	3
Age 8-9	6	2

Table 5.10 Comparing the values of the answers to Q 3 and Q4

In posttest question 3, there were 5 green grasses and 4 brown grasses, so the correct answer was 5.6. In question 4, there are 3 green grasses and 1 brown grass, so the correct answer was 7.5. There were more green grasses in question 3, but the answer to question 3 had a smaller value than question 4. When presented with the counter suggestion, only one participant in the group of 4 to 5 year olds changed her answer. Three participants in the group of 4 to 5 year olds, two participants in the group of 6 to 7 year olds, and three participants in the group of 8 to 9 year olds explained their answer by stating how the number of grasses differed. Three participants in the group of 6 to 7 year olds and four participants in the group of 8 to 9 year olds responded by explaining that the difference between the numbers of the two types of grasses were not the same in the two questions. No participant in the group of 4 to 5 year olds used the difference of two types of grasses in their explanation.

According to Noelting (1980), comparison between 5:4 and 3:1 was a stage IIIA problem. (Ratios with two corresponding terms is a multiple of one another.) Noelting defined the age of accession for a stage as the age where 50% of the participants solved at least one item of the stage. It was found that the age of accession for stage IIIA was 12 years 2 months. In the current study, 62.5% of the participants in the groups of 4 to 5 year olds and 6 to 7 year olds, and 75% of participants in the group of 8 to 9 year olds were capable of providing a greater answer value for 3:1 than 5:3. The difference in the age of accession in the two studies was quite large. One of the main reasons behind this difference is that Noelting counted an item as correct only when the choice was correct and the explanation indicated that the participant had not just compared two terms additively. In the current study, only the answers to the two questions were compared. It was reasonable for Noelting to consider both the explanation and the choice made, since their task required the children to choose one answer from two choices. If only the given options were considered, the chance for a child to get the correct answer by chance was 50%, so this behavioural measure could not accurately reflect the children's understanding if used on its own. On the other hand, participants were required to evaluate the probability of each question individually in this task. The behavioural measure of the current probability estimation task could reflect participants' knowledge more accurately than the choice task. More importantly, the belief that the explicit explanation can completely reflect children's understanding is challenged by the current experiment. Using the explicit explanation to decide whether a behavioural success should be counted as correct may lead to underestimation of children's ability, particularly if the development occurs from the bottom-up direction.

Mix (2002, p. 72) suggested that tasks that do not require an explicit comparison across sets are easier than tasks that do. Noelting's task required explicit comparison. However, for the current experiment, the primary problem just involved evaluation of probability for one set of green grasses and brown grasses. The difference in the nature of the task may also explain the difference between the performances found in the two studies.

5.4 Comparison of the behavioural performance and the performance that demonstrated explicit understanding

5.4.1 Accuracy score of explanation type I

An explanation was classified as type I (Implicit) if the participants said that they did not know, or their explanations did not provide any computation method, and they neither commented on the quantity of grasses, nor made any comparison. If probability estimation is only sustained by explicitly represented knowledge, then for the trials in which participants provided explanation I, there should be no difference in behavioural performance among the different age groups. In the trials in which explanation I was provided, three t-tests were carried out to test whether the accuracy score of the three age groups was the same. The familywise alpha was controlled at .05, therefore alpha for each test was .05/3 = .0167.

In the trials in which explanation I was provided, the mean of accuracy score for the group of 4 to 5 year olds ($M = 2.16$, $SD = 1.66$) was significantly different from the group 6 to 7 year olds ($M = .96$; $SD = 1.14$): $t(119.1) = 4.75$, $p < .01$ (two-tailed). The mean difference between the group of 4 to 5 year olds and the group of 8 to 9 year olds ($M = .65$; $SD = .56$) was also significant: $t(13.95) = 5.16$, $p < .01$. However, the difference between the group of 6 to 7 year olds and the group of 8 to 9 year olds was not significant: $t(50) = .66$, $p > .0167$.

The null hypothesis that the accuracy score of explanation I trials of all three age groups were the same was rejected. Because behavioural performance improved with age even when the participant failed to provide an explicit explanation, it is difficult to explain if explicit knowledge replaced implicit knowledge in the process of development. The behavioural improvement among these trials should be caused by improvement of intuitive estimation, which was sustained by implicit representation. The computation process of intuitive estimation is not available for verbal report and conscious access.

5.4.2 Changes in the pretest and posttest performance of the group of 4 to 5 year olds

In section 5.2.1.2, it was found that the posttest accuracy scores of the group of 4 to 5 year olds was significantly improved when compared with their pretest accuracy scores. Since the participants did not receive any explicit instruction about how to carry out the computation, this improvement should be a result of experience gained in the training task. There was behavioural improvement in accuracy, but was there improvement in the performance of explicit understanding?

	Pretest	Posttest	Training
I	21	24	33
L	-	-	8
C	10	12	17
P	-	-	-
H	-	-	-
F	-	-	-
O	9	4	8

Table 5.11 Explanation distribution of the group of 4 to 5 year olds

It can be seen that, in both pretest and posttest, participants were restricted to using only explanation I, C, and O. None of the three explanations contained any explicit computation procedure that could lead to the correct answer, and nor could these explanations specify a range for the correct answer. At the end of posttest, when the experimenter explicitly asked participants for their method of knowing the answer, none of the participants provided any computational procedure. Therefore, the behavioural improvement was the result of improved intuitive estimation, not the discovery of a correct explicit computational procedure, and this intuitive estimation should be sustained by implicit representation.

5.4.3 Comparing the behavioural adherence of half rule with the performance of explanation type H

Explanation type H required the half rule to be mentioned explicitly in the explanation, which was sustained by the E2/3 representation according to the RR model. If behavioural adherence to the half rule also relied on E2/3 representation, the developmental trend should be similar to the verbal version.

The number of explanation type H provided by each participant among all of the trials in the experiment was counted, and ANOVA of the count was implemented. The between-subjects factors were age group (Age groups 4-5, 6-7, and 8-9) and gender (Male and Female).

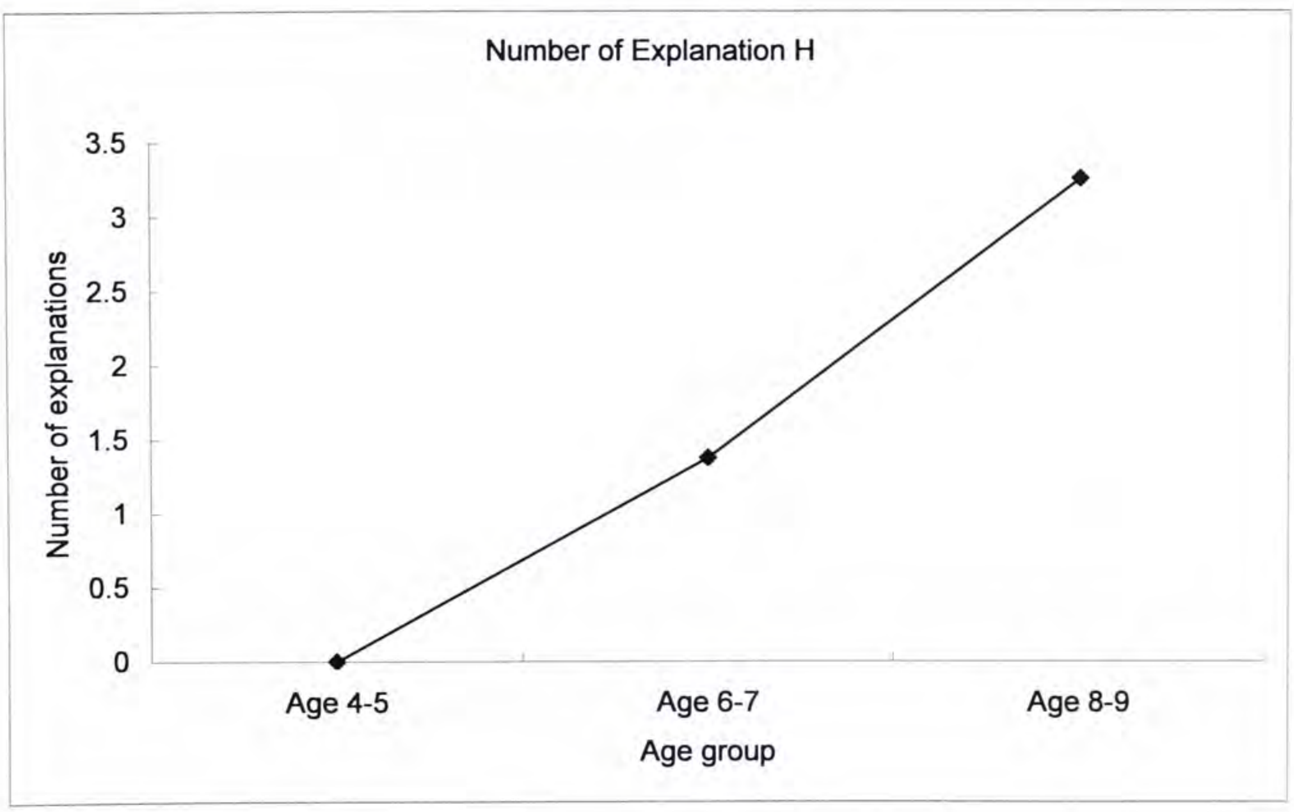

Fig. 5.5 Distribution of explanation type H

The effect of age group was significant, $F(2, 18) = 9.764$, $p < .01$, Partial Eta Squared = .520. The effect of gender, however, was not significant, $F(1, 18) = .146$, $p > .05$. The interaction between age group and gender was also not significant, $F(2, 18) = .506$, $p > .05$. Post hoc test using the Tukey HSD procedure indicated that the only significant performance difference was between the group of 4 to 5 year olds and the group of 8 to 9 year olds. The post hoc LSD (least significant difference) test, which is more able to detect difference with greater power, also agreed with the Tukey HSD test that the difference between the groups of 4 to 5 year olds and 6 to 7 year olds was not significant.

We have shown that the BHalf score of the group of 4 to 5 year olds was significantly different from the group of 6 to 7 year olds and the group of 8 to 9 year olds. If both behavioural and explanation performance were sustained by the same E2/3 representation, it is difficult to explain the significant difference that occurred between the group of 4 to 5 year olds and the group of 6 to 7 year olds. The difference could be explained if explanation type H was sustained by E2/3 representation, and behavioural adherence to the half rule was not sustained by E2/3 representation, but was a representation of a lower-level, such as level-E1 or level-I. This explanation would allow the development of different levels of representation at different speeds.

The results indicated that, in the group of 6 to 7 year olds, the children's behavioural performance was better than the youngest age group, but their performance in giving explanations was similar to the youngest group. Their behavioural development preceded the development of explicit knowledge, so this seems to be an example of the bottom-up direction of development.

5.5 Summary

5.5.1 List of important findings related to the RR model

Regarding to the distinction of implicit and explicit representation, three differences between behavioural performance and performance involving explicit knowledge were found in the current study:

1. Among the trials in which explanation I were provided, the accuracy of estimation improved with age.
2. A different pace of development was found between behavioural adherence to the half rule and explicit explanations of the half rule.
3. In the group of 4 to 5 year olds, the posttest improvement in accuracy was not accompanied by the discovery of more appropriate computational procedures.

The difference in explicit knowledge (which sustains verbal explanation) and implicit knowledge (which sustains intuitive estimation) was observed in these results, and the difference in representation level is a possible explanation for all of these differences.

The following two results suggest the existence of learning from the bottom-up direction in the experiment:

1. In the posttest, accuracy of the group of 4 to 5 year olds improved, but no advanced computation strategy was discovered. This suggests that the implicit knowledge developed into a more advanced stage earlier than the explicit knowledge.
2. The explanations that participants provided for the general questions were less advanced than those in the trials. This should not be the case if the developmental direction was top-down, because general knowledge should be the starting point of development. The result suggests that the participants might have relied on the problem solving process itself for more advanced explanations, which lends support to a bottom-up direction of development.

The U-shaped performance curve among age groups was not observed in the current study. This is not surprising because no previous research found U-shaped curves for the probability task, and the RR model does not predict that the U-shaped performance curve must occur. The issue of whether an intraparticipant U-shaped curve exists was not investigated, because the middle part of the experiment was a training phase. In the training phase correct answers were provided, so it is unfair to compare performance of the training phase with the pretest and posttest phases.

The developmental pattern of more advanced explanations (explanation H and F) were all discovered in trials in which the number of green and brown grasses were equal. After the first occurrence of a more advanced explanation, participants still used less advanced explanations in later trials, until the less advanced explanations were gradually phased out.

5.5.2 List of important findings related to the age-related development of the probability-estimation task

The development of explicit understanding of the probability task was shown by the finding that the most correct explanation that the group of 4 to 5 year olds could provide was explanation C. This reflects that they had some awareness that the ratio of intended and unintended outcome could both affect probability. However, they did not use this knowledge to respond to general questions about computation at the end of the experiment. Also, proper quantification of this awareness had not been developed yet. No participant described the computation method or the steps involved in the computation.

In the group of 6 to 7 year olds, explanation C became the most commonly used explanation and was found in 34.7% of trials. Explanation P, which contains explicit descriptions of computational steps, first appeared in this age group and was found in 21.5% of the trials. Explanation H also started to appear in this age group, but it was used in a very small number of trials (7.6%). At the end of the experiment, all of these participants correctly reported that they had considered both type of grasses. Their responses to the general questions about computation were much better than the group of 4 to 5 year olds, because only two of the participants responded with explanation I.

Explanations involving the concept of fractions or division first appeared in the group of 8 to 9 year olds. However, the most commonly used explanation of this age group was still explanation C (57.6%). Explanation H was found in 18.1% of all trials, which was the second most popular explanation. Both explanation C and explanation H were used by all of the participants. Explanation P was found in 9.7% of all trials, and all of them were subtype 2. In response to general questions, four of the participants mentioned explanation H, which they had used in previous trials.

5.5.3 Responding to key research questions in a nutshell

Like the case of the block-balancing task, differences between behavioural performance and performance that demonstrates explicit understanding was observed in this probability estimation task. The differences were likely to be caused by the difference in explicit representation and implicit representation. There is also evidence to suggest the existence of the bottom-up direction of development. Similar to the findings of the block-balancing experiment of this study, a U-shaped performance curve could not be observed in this experiment. As mentioned before, this did not pose any threat to the plausibility of the RR model. The results of experiment two echo the RR model's description of development.

CHAPTER SIX: CONCLUSION

6.1 Summary of findings in Experiments One and Two

6.1.1 Performance difference that reflects the distinction between implicit and explicit representations

In both experiments, the behavioural performance and the performance that demonstrates explicit understanding were first analyzed individually, and then the performances were compared with each other to see whether there was a difference between them. If there is only one type of representation, then the explicit understanding and the related behavioural performance should have a similar performance level, be affected by the same set of factors, and develop at the same rate. Both Study One and Study Two found the opposite because differences were found between the two types of performance in both tasks. It seems that it is more reasonable to take the position that the two types of performance are sustained by different levels of representation than the same level of representation.

In the block-balancing experiment, the factors that had a significant effect on the Success Score and the Prediction Task were different. For the Success Score, age was a significant main effect, but this was not the case for the Prediction Task. The developmental trends of the two were also different. For the asymmetrical blocks, the

Success Score increased with age. However, this was not the case for the Prediction Task. The performance for the asymmetrical blocks in the Prediction Task was similar in the three age groups, and they all predicted a similar amount of failures for the asymmetrical blocks.

The block-balancing task also provides evidence for the existence of E1 representation, which is a unique invention of the RR model that provides an intermediate representation level along the explicit-implicit dimension. It was found that there was a stage in which the geometric-centre theory could not be verbalized, but could lead to persistent central-area placement and incorrect judgment in the Prediction Task.

In the block-balancing task, examples for all three levels of representation (level-I, level-E1, and level-E2/3) were found. The behavioural skills involved in successfully balancing the blocks was represented at level-I. When the geometric-centre theory was neither available to consciousness nor ready for verbalization, it exerted influence in the form of persistent central-area placement and influenced the prediction result, so the geometric-centre theory was represented at the E1 level at this stage. In later a stage, when the geometric-centre theory was verbalized and available for conscious inference, it was represented at level-E2/3. The discovery of the naïve version of law of torque, in this study, seems to have involved the bottom-up direction of development. The possible developmental pattern may be that after the geometric-centre theory was discovered and represented at level-E2/3, it was compiled into a lower level which sustained the behavioural pattern that reflected the concept. The compilation involved in this suggested process is a kind of top-down development. The three entities mentioned above, namely, the behavioural skills for balancing, the geometric-centre theory, and the naïve version of the law of torque, might coexist in the same person simultaneously. This might explain why some participants could balance the asymmetrical blocks and provide explanation T in Phases One and Two, but predicted that the blocks could not be balanced in the Prediction Task.

In the probability task, the difference between behavioural performance and performance that demonstrates explicit understanding was also observed. First, it is noteworthy that accuracy improved with age, even among the trials in which explanation I was provided. Second, the pace of development was different between behavioural adherence and verbalization of the half rule. It should also be noted that the accuracy improvement found in the group of 4 to 5 year olds was not accompanied by the discovery of more advanced computational procedures.

In the block-balancing task, the different concepts could be clearly classified into the three different levels of the RR model. However, in the probability task, knowledge about probability could only be roughly classified into implicit and explicit representation. This is not a deficit of the RR Model; the rough classification is only due to insufficient evidence to judge whether knowledge that cannot be verbalized should belong to level-I or level-E1 representation. Intuitive estimation ability was sustained by implicit representation, and the verbal explanations given by the participants were different explicit concepts. The implicit representation and explicit representation coexisted in the same person's mind simultaneously. Although older participants had more advanced explicit knowledge about probability, when they provided explanation I in a trial, they made intuitive estimations in that trial. If explicit representation replaced implicit representation in the developmental process, implicit representation was no longer found in older participants, because their implicit representation had already been replaced by explicit representation and their accuracy should not be better than the younger participants. However, Study Two found that among the trials in which explanation I was provided, the Accuracy Score still improved with age. This shows that the implicit representation was not replaced in the developmental process.

6.1.2 Top-down or bottom-up learning

Examples for top-down and bottom-up development were found in both experiments. In the block-balancing experiment, it appeared that correct explanations were not the prerequisite for behavioural success in block-balancing. On the other hand, in the trials in which correct explanations were provided, all of the block-balancing actions were successful. Furthermore, 6 participants were able to balance asymmetrical blocks behaviourally, but could not predict this success in the Prediction Task. This suggests that behavioural improvement precedes improvement in explicit knowledge, which is evidence for the bottom-up direction of development. An example of top-down development was found in the second stage development of the naïve version of the law of torque, in which explicit knowledge appeared to be the prerequisite of the behavioural pattern that reflected this knowledge.

In the probability-estimation experiment, it was found that behavioural adherence to the half rule improved earlier than the explicit verbalization of the half rule. It was also found that participants' in-trial explanations were generally more advanced than their responses to general questions. Moreover, the youngest age group improved in behavioural performance before they discovered more advanced computation methods. All of these findings are evidence to suggest a bottom-up direction of development for the probability-estimation task. An example of top-down development was found in one participant, CF1, who could apply the concept of fractions when she tried to solve the probability problem. She told the experimenter that her mother had taught her explicit knowledge about fractions.

6.1.3 Is there a U-shaped performance curve?

A U-shaped performance curve across age groups was not found in the block-balancing task, nor in the probability-estimation task. As indicated in the literature review, Karmiloff-Smith (1992) did not say that the U-shaped performance curve is an essential for the RR model. Quoted from Karmiloff-Smith (1992, p. 20): "The temporary disregard for features of the external environment during phase 2 can lead to new errors and inflexibilities. *This can, but does not necessarily* [italics added], give rise to decrease in successful behaviour—a U-shaped performance curve." Therefore, an absence of the U-shaped curve does not pose any threat to the basic postulates of the RR model. Given the abundant evidence for differences in performance between implicit and explicit representation, and the evidence for bottom-up learning, this study lends support to the basic postulates of the RR model.

6.1.4 Other findings

In both experiments, it was found that, once discovered, the more advanced explanations (explanation T in the block-balancing task, and explanations F and H in the probability estimation task) were not applied to all later trials. Siegler and Jenkins (1989) found that the repeated use of a more advanced strategy was related to the participants' level of insight into that strategy. If the participants were not aware that a certain strategy was more advanced, they were less likely to use the newly discovered strategy in later trials.

In the block-balancing task, more advanced explanations were first discovered in the more challenging asymmetrical block trials. This may be due to the fact that less advanced explanations could not account for the success in balancing the asymmetrical blocks. In the probability estimation task, more advanced explanations were first found in the easier trials, in which equal numbers of grasses were involved in the questions. This may be due to the fact that the concept of "half" is easier to master in this type of context, so that the participants were in a better position to understand the relation of "half" and division or fractions. The different findings of the two experiments suggested that the chance of acquiring a more advanced explanation depends on the content of the concept and the nature of the task.

6.1.5 Summary

To summarize, the current study provided empirical findings that favour the RR model. If the distinction between implicit and explicit representation does not exist in the first place, it is unreasonable to suppose that the representation redescription process exists, given that the representation redescription process is a process of turning implicit knowledge into explicit knowledge. Evidence was found to suggest a distinction between implicit and explicit representation, and also supported the existence of a bottom-up direction of development. These are important assumptions of the RR model. Affirmation of these assumptions set the stage for the further investigation of the core of the RR model: the representation redescription process.

The RR model had been challenged by some researchers, like Krist et al. (2005), who claimed that they had neutralized important evidence for the RR model by showing there was no U-shaped performance curve. However, the absence of the U-shaped performance curve does not pose a fatal challenge to the validity of the RR model. The reason behind the existence of the U-shaped performance curve in the block-balancing task is the geometric-centre theory, which captures some commonalities about balancing, but overgeneralization of this theory to the asymmetrical blocks led to a decline in block-balancing performance. The current study proved the existence of the implicit geometric-centre theory. Since the effect of this implicit theory was not limited to the block-balancing behaviour, but also the performance in the Prediction Task, the geometric-centre theory should be represented at level-E1. This provided evidence for the existence of the E1 level and helped the RR model to regain its plausibility. The postulation and validation of the E1 representation level, which breaks the commonly assumed implicit/explicit dichotomy, is a unique contribution of the RR model.

The current study also provided a new piece of evidence for supporting the RR model as a domain-general developmental process. The predictions made by the RR model were confirmed in both the block-balancing experiment and the probability estimation experiment. To the best of the author's knowledge, this is the first time the RR model has been successfully applied in a probability task.

6.2 Implications

The rich findings of the current study strongly suggest that the explicit-implicit dimension is a useful dimension for understanding conceptual development. The owner of a concept may first own the knowledge without awareness of its presence, and the awareness and flexibility of the knowledge increases as a result of development. The opposite may also happen, because loss of conscious control and increase of efficiency may also be the effect of development. By applying the explicit-implicit dimension to other areas, further interesting or insightful findings may be obtained.

The current study has shown that the RR model has good empirical support, and it is a useful model for understanding conceptual development. The application of the model has been extended to the probability domain. The RR model can act as a bridge for integrating the findings of developmental psychology and cognitive psychology, so that both subfields in psychology can be enriched, and their application area and explanation power can be extended.

The findings of the current study also have implications for experimental design. The behavioural performance and verbal explanations should not be viewed as two sides of the same coin. The two may develop at a different pace. In some experiments, correctness of an item does not only require children to make a correct choice, but also requires a correct explanation. If the task involves bottom-up development, this kind of experimental design risks underestimating the children's ability. Other than that, it may not be appropriate to view children's verbal explanation as a complete and accurate reflection of the knowledge that they possess. Even if a child cannot give a correct explanation, it does not mean that the child must be totally ignorant of the topic. The knowledge that the child possesses may be represented at E1 or level-I, which needs to be observed by a means other than a verbal report.

The findings of the present study also have implications for teaching and learning. In this study, it was found that even if the participants could verbally give a correct explanation, this did not mean the participant had finished the learning or the development process. The participant might fail to use the correct knowledge in similar situations, and an incorrect concept might coexist with a correct concept. More importantly, the participants might not understand to what extent their concepts were correct. While students should be encouraged to reflect on their own concepts and their own thinking process, the teacher should help students to identify assumptions that the students might not be aware of, and help them to gain fuller insights into their own concepts and knowledge. This can help students to consolidate their knowledge, and help them to gain greater flexibility in the application of their knowledge.

Findings about the implicit and explicit levels of representation that illustrated conceptual understanding shed light on the methods of instruction. The teachers' role is not limited to presenting well organized, explicit knowledge. In the process of solving a problem, or when students are trying to construct their concepts, they may go through a phase in which their ideas are at a representation level that is unavailable to conscious access or a verbal report. Students cannot be fully aware of what is in their minds, and cannot verbalize their ideas. At this moment, telling students some explicit knowledge may not help much, and may cause them even more confusion. A teacher's role should be that of a facilitator at this phase. The teacher needs to be very sensitive, actively understand what is in the students' minds, and actively organize meaning from the students' seemingly meaningless and unorganized speech. Students may already have a partial understanding of a subject, and further learning will be most effective if the teacher can help the students to be aware of what they know and to build on the ideas that they already have.

Such findings should also help teachers to review their students' methods in the assessment of their learning. A paper-and-pencil test is the most common kind of test that is used for accessing student's understanding. However, for some subjects, this may not be the ideal way to access their conceptual understanding. For example, if a teacher wants to assess whether students know how to diagnose and fix computer problems, it may not be appropriate to use a paper-and-pencil test only. Students may be able to perform the task successfully, but not be able to present their conceptual understanding by using pen and paper. The reverse can also be true, students may be able to remember the text from a textbook, but be unable to apply what they remember when they face a dysfunctional computer. A practical test cannot be replaced by a paper-and-pencil test in a practical subject such as this.

6.3 Limitations

This study aimed to investigate the change of implicit conceptual representation to explicit conceptual representation. To find out how children's representations were changed, it was necessary to analyse their verbal explanations in detail, so trial-by-trial examination of the children's explanations was thus carried out. Combining all of the participants' changes into one age group would lose information about developmental trajectory, so to find out the commonality of development patterns in this study, the changes of each participant were traced individually. Given the limited manpower and the workload involved, the number of participants in each age group could not be too large. This was an unavoidable trade-off between higher temporal clarity and individuality, against the expense of capturing a more generalized picture.

To find out the content of children's explicit representation, it was necessary to ask questions to elicit verbal reports of their understanding. The dilemma that arose from this questioning was that asking more questions elicited more verbal reports and provided more information to help understand what was in the children's minds: however, the questions unavoidably gave hints or sensitized the children's attention to the important features of the concepts. The questions were intended to be a tool for understanding the children's concepts, not the source of change in the experimental design. This balance is difficult to achieve. In this study, questions that went to a greater depth and would have more effect on the children's conceptual understanding were asked at the end of experiment. Although this placement of questions was to avoid an undesired influence on the developmental pattern, this might have scarified a deeper understanding of the children's minds in earlier trials.

In the probability task, it was found that fractions with different numerators and denominators had an effect on the accuracy score. In this experiment, the author had tried her best to use questions of similar types in the pretest and posttest sections. Better control could be achieved if the questions in the pretest and the posttest were similar in accuracy score (i.e., similar in difficulty). To the best of the author's knowledge, the accuracy score obtained by using a similar paradigm was not available. Moreover, this control could be quite difficult to achieve across age groups, because the difficulty of the questions varied with age, which is exemplified by the finding that the best and worst questions were different among the age groups in this study.

Due to the limited resource involved, only one school having both primary and kindergarten sections was invited to participate in the experiment. Most of the students in this school have a middle-class background, so the results may not be applicable to another population.

6.4 Suggestions for further studies

The RR model is intended to be a domain-general model. Karmiloff-Smith (1992) had applied the RR model to the domains of language, physics, mathematics, theory of mind, and notation in her book *Beyond Modularity*. To the best of the author's knowledge, this is the first time that the RR model has been applied to the study of the concept of probability, and this application has brought meaningful findings in the current study. It is suggested that the RR model could be further applied to the study of conceptual development in new domains. For instance, the model could be applied to study the development of the concept of "justice". It is observed that many people have a very strong sense of "just" versus "unjust," but many of them cannot articulate their own theory of justice. Tasks could be designed to obtain children's behavioural performance with respect to the concept of justice, such as by making choices between several alternatives. Verbal explanations should also be obtained in the tasks, so that the children's explicit understanding could be analyzed.

In the probability-estimation task of the current study, it is difficult to tell whether the implicit knowledge was represented at level-I or the E1-level. A new experiment could be designed to further classify this knowledge into a more precise representation type.

For both the block-balancing task and the probability-estimation task, the age groups involved may be further expanded to gain a more complete picture of cross-age development. For the block-balancing task, younger participants may be included to investigate when the geometric-centre theory begins to emerge. An experiment could be designed, for a probability-estimation task, to obtain an accuracy score that is free of the practice effect, by randomizing the questions, and by the removal of feedback in the form of correct answers.

REFERENCES

Acredolo, C., O'Connor, J., Banks, L, & Horobin, K. (1989). Children's ability to make probability estimates: Skills revealed through application of Anderson's functional measurement methodology. *Child Development*, *60*, 933-945.

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology*, *25*, 468-573.

Anderson, H. H. (1980). Information integration theory in developmental psychology. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children* (pp. 1-45). Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369-406.

Antell, E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, *54*, 695-701.

Baillargeon, R., & Hanko-Summers, S. (1990). Is the top object adequately supported by the bottom object? Young infants' understanding of support relations. *Cognitive Development*, *5*, 29-53.

Brooks, D. N., & Baddeley, A. (1976). What can amnesic patients learn? *Neuropsychologia*, *14*, 111-122.

Bryant, P. E. (1974). *Perception and understanding in young children.* London: Methuen.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 259-292). Hillsdale, NJ: Lawrence Erlbaum Associates.

Case, R. (1985). *Intellectual development: A systematic reinterpretation.* New York: Academic Press.

Chapman, R. H. (1975). The development of children's understanding of proportions. *Child Development*, *36*, 779-788.

Chomsky, N. (1980). *Rules and representations.* New York: Columbia University Press.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377-395.

Cohen, N. J. (1984). Preserved learning capacity in amnesia: Evidence for multiple memory systems. In L. R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (pp. 83-103). New York: Guilford.

Cohen, N. J., Eichenbaum, H., Deacedo, B. S., & Corkein, S. (1985). Different memory systems underlying acquisition of procedural and declarative knowledge. In D. S. Olton, E. Gamzu, & S. Corkin (Eds.), *Memory dysfunctions: An integration of animal and human research from preclinical and clinical perspective* (pp. 54-71). New York: Annals of the New York Academy of Sciences.

Cohen, N., & Squire, L. R. (1980). Preserved learning and retention of pattern analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*, 207-209.

Critten, S., Pine, K., & Steffler, D. (2007). Spelling development in young children: A case of representational redescription? *Journal of Education Psychology*, *99*, 207-220.

Daum, I., Channon, S., & Canavar, A. (1989). Classical conditioning in patients with severe memory problems. *Journal of Neurology, Neurosurgery & Psychiatry*, *52*, 47-51.

Dienes, Z., Altmann, G. T. M., Kwan L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 875-878.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Science*, *22*, 735-808.

Dixon, J. A., & Bangert, A. S. (2002). The prehistory of discovery: Precursors of representation change in solving gear system problem. *Developmental Psychology*, *38*, 918-933.

Dixon, J. A., & Dohn, M. (2003). Redescription disembeds relations: Evidence from relational transfer and use in problem solving. *Memory & Cognition*, *31*, 1082-1093.

Dixon, J. A., & Kelley, E. (2007). Theory revision and redescription: Complementary process in knowledge acquisition. *Current Directions in Psychological Science*, *16*, 111-115.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, *87*, 477-531.

Fodor, J. (1975). *The language of thought.* New York: Thomas Y. Crowell.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Gelman, R., & Meck, E. (1986). The notion of principle: The case of counting. In J. Hiebert (Ed.), *The relationship between procedural and conceptual competence* (pp. 29-57). Hillsdale, NJ: Lawrence Erlbaum Associates.

Glisky, E. L., Schacter, D. L., & Tulving, E. (1986). Computer learning by memory-impaired patients: Acquisition and retention of complex knowledge. *Neuropsychologia*, *24*, 313-328.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories.* Cambridge, MA: MIT Press.

Jackendoff, R. (1989). What is a concept, that a person may grasp it? *Mind and Language*, *4*, 68-102.

Karmiloff-Smith, A. (1979a). Micro- and macro-developmental change in language acquisition and other representational systems. *Cognitive Science*, *3*, 81-118.

Karmiloff-Smith, A. (1979b). *A functional approach to child language.* Cambridge, England: Cambridge University Press.

Karmiloff-Smith, A. (1984). Children's problem solving. In M. E. Lamb, A. L. Brown, & B. Rogoff (Eds.), *Advances in developmental psychology* (Vol. 3, pp. 39-90). Hillsdale, NJ: Lawrence Erlbaum Associates.

Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition*, *23*, 95-147.

Karmiloff-Smith, A. (1992). *Beyond modularity.* Cambridge, MA: MIT Press.

Karmiloff-Smith, A., & Inhelder, B. (1974/75). If you want to get ahead, get a theory. *Cognition*, *3*, 195-212.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Krist, H., Horz, H., & Schönfeld, T. (2005). Children's block balancing revisited: No evidence for representational redescription. *Swiss Journal of Psychology*, *63*, 183-193.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, *60* (3, Serial No. 245).

Landis, R. J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Locke, J. (1975). *An essay concerning human understanding.* (P. H. Nidditch, Ed.) Oxford, England: Oxford University Press.

Lovett, S. B., & Singer, J. A. (1991, April). The development of children's understanding of probability: Perceptual and quantitative conceptions. Paper presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA.

Meguro, K., Senaha, M., Caramelli, P., Ishizaki, J. Chubacci, R., Meguro, M., Ambo, H., Nitrini, R., & Yamadori, A. (2003). Language deterioration in four Japanese-Portuguese bilingual patients with Alzheimer's disease: A transcultural study of Japanese elderly immigrants in Brazil. *Psychogeriatrics*, *3*, 63-68.

Messer, D., Joiner, R., Light, P., & Littleton, K. (1998). The development of children's ability to balance objects on a fulcrum: Examining Karmiloff-Smith's model of cognitive development. *Educational Psychology*, *18*, 81-96.

Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology*, *69*, 199-221.

Mix, K. S. (2002). *Quantitative development in infancy and early childhood.* New York: Oxford University Press.

Moore, C. F., Dixon, F. A., & Haines, B. A. (1991). Components of understanding in proportional reasoning: A fuzzy set representation of developmental progressions. *Child Development*, *62*, 441-459.

Noelting, G. (1980). The development of proportional reasoning and the ratio concept: Part I – Differentiation of stages. *Educational Studies in Mathematics*, *11*, 217-253.

Nunes, T., & Bryant, P. (1996). *Children doing mathematics*. Oxford, England: Blackwell.

Nunes, T., Bryant, P., & Bindman, M. (1997). Morphological spelling strategies: Developmental stages and process. *Development Psychology*, *33*, 637-649.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism.* Amsterdam: John Benjamins.

Paradis, M., & Gopnik, M. (1997). Compensatory strategies in familial language impairment. *Journal of Neurolinguistics*, *10*, 173-185.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Piaget, J. (1955). *The child's construction of reality.* London: Routledge & Kegan Paul.

Piaget, J., & Inhelder, B. (1975) *The origin of the idea of chance in children* (L. Leake, P. Burrell & H. D. Fishbein, trans.). London: Routledge & Kegan Paul. (Original work published in 1951.)

Pine, K. J., Lufkin, N., & Messer, D. (2004). More gesture than answer: Children learning about balance. *Developmental Psychology, 40*, 1059-1067.

Pine, K. J., & Messer, D. (1998). Group collaboration effects and the explicitness of children's knowledge. *Cognitive Development, 13*, 109-126.

Pine, K. J., & Messer, D. (1999). What children do and what children know: Looking beyond success using Karmiloff-Smith's RR framework. *New Ideas in Psychology, 17*, 17-30.

Pine, K. J., & Messer, D. (2000). The effect of explaining another's action on children's implicit theories of balance. *Cognition and Instruction, 18*, 35-51.

Pine, K. J., & Messer, D. (2003). The development of representations as children learn about balancing. *British Journal of Developmental Psychology, 21*, 285-301.

Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis.* Cambridge, MA: MIT Press.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855-863.

Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: A review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 239-272). Chichester, England: Wiley.

Ross, B. M., & Hoemann, H. W. (1975). The attainment of formal operations: A comparison of probability concepts in deaf and hearing adolescents. *Genetic Psychology Monographs*, *91*, 61-119.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 501-518.

Schacter, D. L. (1999). Implicit vs. explicit memory. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 394-395). Cambridge, MA: MIT Press.

Schacter, D. L., & Graf, P. (1986). Preserved learning in amnesic patients: Perspectives from research on direct priming. *Journal of Clinical Experimental Neuropsychology*, *6*, 727-743.

Shiu, L. P., & Sin, C.-Y. (2006). Top-down, middle-out, and bottom-up processes: A cognitive perspective of teaching and learning economics. *International Review of Economic Education*, *5*, 60-72.

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, *127*, 366-397.

Snowden, J. S., Neary, D., & Mann, D. M. A. (1996). *Fronto-Temporal lobar degeneration: Fronto-Temporal dementia, progressive aphasia, semantic dementia.* New York: Churchill Livingstone.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29-56.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605-632.

Spinillo, G. H., & Bryant, P. (1991). Children's proportional judgements: The importance of "Half." *Child Development*, *62*, 427-440.

Spinillo, G. H., & Bryant, P. (1999). Proportional reasoning in young children: Part-Part comparisons about continuous and discontinuous quantity. *Mathematical Cognition*, 5, 181-197.

Squire, L. R. (1982). The neuropsychology of human memory. *Annual Review of Neuroscience*, *5*, 241-273.

Squire, L. R. (1994). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 203-232). Cambridge, MA: MIT Press.

Squire, L. R., & Frambach, M. (1990). Cognitive skill learning in amnesia. *Psychobiology*, *18*, 109-117.

Steffler, D. J. (2001). Implicit cognition and spelling development. *Developmental Review*, *21*, 168-204.

Vygotsky, L. S. (1934/1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky: Vol 1. Problems of general psychology* (pp. 39-285). New York: Plenum Press.